B33

PCT WORLD INTELLECTUAL PROPERTY ORGANIZATION
International Bureau

INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(51) International Patent Classification 6 : C07H 21/00 — A1

(11) International Publication Number: WO 99/14228

(43) International Publication Date: 25 March 1999 (25.03.99)

(21) International Application Number: PCT/US98/19325

(22) International Filing Date: 16 September 1998 (16.09.98)

(30) Priority Data:
60/059,304 17 September 1997 (17.09.97) US
60/066,172 18 November 1997 (18.11.97) US

(63) Related by Continuation (CON) or Continuation-in-Part (CIP) to Earlier Applications
US 60/059,304 (CIP)
Filed on 17 September 1997 (17.09.97)
US 60/066,172 (CIP)
Filed on 18 November 1997 (18.11.97)

(71) Applicant *(for all designated States except US)*: AFFYMETRIX, INC. [US/US]; 3380 Central Expressway, Santa Clara, CA 95051 (US).

(72) Inventors; and

(75) Inventors/Applicants *(for US only)*: LIPSHUTZ, Robert, J. [US/US]; 970 Palo Alto Avenue, Palo Alto, CA 94301 (US). CHEE, Mark [AU/US]; 3199 Waverley Street, Palo Alto, CA 94306 (US). FAN, Jian-Bing [CN/US]; Apartment 20, 275 Ventura Avenue, Palo Alto, CA 94306 (US). BERNO, Anthony [CA/US]; 570 South 12th Street, San Jose, CA 95112 (US).

(74) Agents: LIEBESCHUETZ, Joe et al.; Townsend and Townsend and Crew LLP, 8th floor, Two Embarcadero Center, San Francisco, CA 94111-3834 (US).

(81) Designated States: AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, CA, CH, CN, CU, CZ, DE, DK, EE, ES, FI, GB, GE, GH, GM, HR, HU, ID, IL, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MD, MG, MK, MN, MW, MX, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TR, TT, UA, UG, US, UZ, VN, YU, ZW, ARIPO patent (GH, GM, KE, LS, MW, SD, SZ, UG, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GW, ML, MR, NE, SN, TD, TG).

Published
With international search report.
Before the expiration of the time limit for amending the claims and to be republished in the event of the receipt of amendments.

(54) Title: GENETIC COMPOSITIONS AND METHODS

(57) Abstract

The invention provides nucleic acid segments of the human genome including polymorphic sites. Allele-specific primers and probes hybridizing to regions flanking these sites are also provided. The nucleic acids, primers and probes are used in applications such as forensics, paternity testing, medicine and genetic analysis.

FOR THE PURPOSES OF INFORMATION ONLY

Codes used to identify States party to the PCT on the front pages of pamphlets publishing international applications under the PCT.

AL	Albania	ES	Spain	LS	Lesotho	SI	Slovenia
AM	Armenia	FI	Finland	LT	Lithuania	SK	Slovakia
AT	Austria	FR	France	LU	Luxembourg	SN	Senegal
AU	Australia	GA	Gabon	LV	Latvia	SZ	Swaziland
AZ	Azerbaijan	GB	United Kingdom	MC	Monaco	TD	Chad
BA	Bosnia and Herzegovina	GE	Georgia	MD	Republic of Moldova	TG	Togo
BB	Barbados	GH	Ghana	MG	Madagascar	TJ	Tajikistan
BE	Belgium	GN	Guinea	MK	The former Yugoslav Republic of Macedonia	TM	Turkmenistan
BF	Burkina Faso	GR	Greece			TR	Turkey
BG	Bulgaria	HU	Hungary	ML	Mali	TT	Trinidad and Tobago
BJ	Benin	IE	Ireland	MN	Mongolia	UA	Ukraine
BR	Brazil	IL	Israel	MR	Mauritania	UG	Uganda
BY	Belarus	IS	Iceland	MW	Malawi	US	United States of America
CA	Canada	IT	Italy	MX	Mexico	UZ	Uzbekistan
CF	Central African Republic	JP	Japan	NE	Niger	VN	Viet Nam
CG	Congo	KE	Kenya	NL	Netherlands	YU	Yugoslavia
CH	Switzerland	KG	Kyrgyzstan	NO	Norway	ZW	Zimbabwe
CI	Côte d'Ivoire	KP	Democratic People's Republic of Korea	NZ	New Zealand		
CM	Cameroon			PL	Poland		
CN	China	KR	Republic of Korea	PT	Portugal		
CU	Cuba	KZ	Kazakstan	RO	Romania		
CZ	Czech Republic	LC	Saint Lucia	RU	Russian Federation		
DE	Germany	LI	Liechtenstein	SD	Sudan		
DK	Denmark	LK	Sri Lanka	SE	Sweden		
EE	Estonia	LR	Liberia	SG	Singapore		

GENETIC COMPOSITIONS AND METHODS

BACKGROUND OF THE INVENTION

The genomes of all organisms undergo spontaneous mutation in the course of their continuing evolution generating variant forms of progenitor sequences (Gusella, *Ann. Rev. Biochem.* 55, 831-854 (1986)). The variant form may confer an evolutionary advantage or disadvantage relative to a progenitor form or may be neutral. In some instances, a variant form confers a lethal disadvantage and is not transmitted to subsequent generations of the organism. In other instances, a variant form confers an evolutionary advantage to the species and is eventually incorporated into the DNA of many or most members of the species and effectively becomes the progenitor form. In many instances, both progenitor and variant form(s) survive and co-exist in a species population. The coexistence of multiple forms of a sequence gives rise to polymorphisms.

Several different types of polymorphism have been reported. A restriction fragment length polymorphism (RFLP) means a variation in DNA sequence that alters the length of a restriction fragment as described in Botstein et al., *Am. J. Hum. Genet.* 32, 314-331 (1980). The restriction fragment length polymorphism may create or delete a restriction site, thus changing the length of the restriction fragment. RFLPs have been widely used in human and animal genetic analyses (see WO 90/13668; W090/11369; Donis-Keller, *Cell* 51, 319-337 (1987); Lander et al., *Genetics* 121, 85-99 (1989)). When a heritable trait can be linked to a particular RFLP, the

presence of the RFLP in an individual can be used to predict the likelihood that the animal will also exhibit the trait.

Other polymorphisms take the form of short tandem repeats (STRs) that include tandem di-, tri- and tetra-nucleotide repeated motifs. These tandem repeats are also referred to as variable number tandem repeat (VNTR) polymorphisms. VNTRs have been used in identity and paternity analysis (US 5,075,217; Armour et al., *FEBS Lett.* 307, 113-115 (1992); Horn et al., WO 91/14003; Jeffreys, EP 370,719), and in a large number of genetic mapping studies.

Other polymorphisms take the form of single nucleotide variations between individuals of the same species. Such polymorphisms are far more frequent than RFLPs, STRs and VNTRs. Some single nucleotide polymorphisms occur in protein-coding sequences, in which case, one of the polymorphic forms may give rise to the expression of a defective or other variant protein and, potentially, a genetic disease. Examples of genes, in which polymorphisms within coding sequences give rise to genetic disease include β-globin (sickle cell anemia) and CFTR (cystic fibrosis). Other single nucleotide polymorphisms occur in noncoding regions. Some of these polymorphisms may also result in defective protein expression (e.g., as a result of defective splicing). Other single nucleotide polymorphisms have no phenotypic effects.

Single nucleotide polymorphisms can be used in the same manner as RFLPs, and VNTRs but offer several advantages. Single nucleotide polymorphisms occur with greater frequency and are spaced more uniformly throughout the genome than other forms of polymorphism. The greater frequency and uniformity of single nucleotide polymorphisms means that there is a greater probability that such a polymorphism will be found in close proximity to a genetic locus of interest than would be the case for other polymorphisms. Also, the different forms of characterized single nucleotide polymorphisms are often easier to distinguish that other types of polymorphism (e.g., by use of assays employing allele-specific hybridization probes or primers).

Despite the increased amount of nucleotide sequence data being generated in recent years, only a minute proportion of the total repository of polymorphisms in humans and other organisms has so far been identified. The paucity of polymorphisms hitherto identified is due to the large amount of work required for their detection by conventional methods. For example, a conventional approach to identifying polymorphisms might be to sequence the same stretch of oligonucleotides in a population of individuals by dideoxy sequencing. In this type of approach, the amount of work increases in proportion to both the length of sequence and the number of individuals in a population and becomes impractical for large stretches of DNA or large numbers of persons.

SUMMARY OF THE CLAIMED INVENTION

The invention provides nucleic acid segments of between 10 and 100 bases from a fragment shown in Table 1, column 1 including a polymorphic site. Complements of these segments are also included. The segments can be DNA or RNA, and can be double- or single-stranded. Some segments are 10-20 or 10-50 bases long. Preferred segments include a diallelic polymorphic site. The base occupying the polymorphic site in the segments can be the reference (Table 1, column 3) or an alternative base (Table 1, column 5).

The invention further provides allele-specific oligonucleotides that hybridizes to a segment of a fragment shown in Table 1, column 8 or its complement. These oligonucleotides can be probes or primers. Also provided are isolated nucleic acids comprising a sequence of Table 1, column 8, or the complement thereto, in which the polymorphic site within the sequence is occupied by a base other than the reference base shown in Table 1, column 3.

The invention further provides a method of analyzing a nucleic acid from an individual. The method determines which base is present at any one of the polymorphic sites shown in Table 1. Optionally, a set of bases occupying a set of the polymorphic sites shown in Table 1 is determined. This type of analysis can be performed on a plurality of individuals who

are tested for the presence of a disease phenotype. The presence or absence of disease phenotype can then be correlated with a base or set of bases present at the polymorphic sites in the individuals tested.

The invention further provides computer-readable storage medium for storing data for access by an application program being executed on a data processing system. Such a medium comprises a data structure stored in the computer-readable storage medium, the data structure including information resident in a database used by the application program. The data structure includes a plurality of records, each record of the plurality comprising information identifying a polymorphisms shown in Table 1.

The invention further provides a signal carrying data for access by an application program being executed on a data processing system. A data structure is encoded in the signal. The data structure includes information resident in a database used by the application program. Such information includes a plurality of records, each record of the plurality comprising information identifying a polymorphism shown in Table 1.

BRIEF DESCRIPTION OF THE FIGURES

Figs. 1A and 1B depict computer systems suitable for storing and transmitting information relating to the polymorphisms of the invention.

DEFINITIONS

An oligonucleotide can be DNA or RNA, and single- or double-stranded. Oligonucleotides can be naturally occurring or synthetic, but are typically prepared by synthetic means. Preferred oligonucleotides of the invention include segments of DNA, or their complements including any one of the polymorphic sites shown in Table 1. The segments are usually between 5 and 100 bases, and often between 5-10, 5-20, 10-20, 10-50, 15-50, 15-100, 20-50 or 20-100 bases. The polymorphic site can occur within any position of the segment. The segments can be from any of the allelic forms of DNA shown in Table 1.

Hybridization probes are oligonucleotides capable of binding in a base-specific manner to a complementary strand of nucleic acid. Such probes include peptide nucleic acids, as described in Nielsen et al., *Science* 254, 1497-1500 (1991).

The term primer refers to a single-stranded oligonucleotide capable of acting as a point of initiation of template-directed DNA synthesis under appropriate conditions (*i.e.*, in the presence of four different nucleoside triphosphates and an agent for polymerization, such as, DNA or RNA polymerase or reverse transcriptase) in an appropriate buffer and at a suitable temperature. The appropriate length of a primer depends on the intended use of the primer but typically ranges from 15 to 30 nucleotides. Short primer molecules generally require cooler temperatures to form sufficiently stable hybrid complexes with the template. A primer need not reflect the exact sequence of the template but must be sufficiently complementary to hybridize with a template. The term primer site refers to the area of the target DNA to which a primer hybridizes. The term primer pair means a set of primers including a 5' upstream primer that hybridizes with the 5' end of the DNA sequence to be amplified and a 3', downstream primer that hybridizes with the complement of the 3' end of the sequence to be amplified.

Linkage describes the tendency of genes, alleles, loci or genetic markers to be inherited together as a result of their location on the same chromosome, and can be measured by percent recombination between the two genes, alleles, loci or genetic markers.

Polymorphism refers to the occurrence of two or more genetically determined alternative sequences or alleles in a population. A polymorphic marker or site is the locus at which divergence occurs. Preferred markers have at least two alleles, each occurring at frequency of greater than 1%, and more preferably greater than 10% or 20% of a selected population. A polymorphic locus may be as small as one base pair. Polymorphic markers include restriction fragment length polymorphisms, variable number of tandem repeats (VNTR's), hypervariable regions, minisatellites, dinucleotide repeats,

trinucleotide repeats, tetranucleotide repeats, simple sequence repeats, and insertion elements such as Alu. The first identified allelic form is arbitrarily designated as a the reference form and other allelic forms are designated as alternative or variant alleles. The allelic form occurring most frequently in a selected population is sometimes referred to as the wildtype form. Diploid organisms may be homozygous or heterozygous for allelic forms. A diallelic polymorphism has two forms. A triallelic polymorphism has three forms.

A single nucleotide polymorphism occurs at a polymorphic site occupied by a single nucleotide, which is the site of variation between allelic sequences. The site is usually preceded by and followed by highly conserved sequences of the allele (e.g., sequences that vary in less than 1/100 or 1/1000 members of the populations).

A single nucleotide polymorphism usually arises due to substitution of one nucleotide for another at the polymorphic site. A transition is the replacement of one purine by another purine or one pyrimidine by another pyrimidine. A transversion is the replacement of a purine by a pyrimidine or vice versa. Single nucleotide polymorphisms can also arise from a deletion of a nucleotide or an insertion of a nucleotide relative to a reference allele.

Hybridizations are usually performed under stringent conditions, for example, at a salt concentration of no more than 1 M and a temperature of at least 25°C. For example, conditions of 5X SSPE (750 mM NaCl, 50 mM NaPhosphate, 5 mM EDTA, pH 7.4) and a temperature of 25-30°C are suitable for allele-specific probe hybridizations.

An isolated nucleic acid means an object species invention that is the predominant species present (*i.e.*, on a molar basis it is more abundant than any other individual species in the composition). Preferably, an isolated nucleic acid comprises at least about 50, 80 or 90 percent (on a molar basis) of all macromolecular species present. Most preferably, the object species is purified to essential homogeneity (contaminant species cannot be detected in the composition by conventional detection methods).

Linkage disequilibrium or allelic association means the preferential association of a particular allele or genetic marker with a specific allele, or genetic marker at a nearby chromosomal location more frequently than expected by chance for any particular allele frequency in the population. For example, if locus X has alleles a and b, which occur equally frequently, and linked locus Y has alleles c and d, which occur equally frequently, one would expect the combination ac to occur with a frequency of 0.25. If ac occurs more frequently, then alleles a and c are in linkage disequilibrium. Linkage disequilibrium may result from natural selection of certain combination of alleles or because an allele has been introduced into a population too recently to have reached equilibrium with linked alleles.

A marker in linkage disequilibrium can be particularly useful in detecting susceptibility to disease (or other phenotype) notwithstanding that the marker does not cause the disease. For example, a marker (X) that is not itself a causative element of a disease, but which is in linkage disequilibrium with a gene (including regulatory sequences) (Y) that is a causative element of a phenotype, can be used detected to indicate susceptibility to the disease in circumstances in which the gene Y may not have been identified or may not be readily detectable.

The present invention includes the use of any of the polymorphic forms shown in Table 1 as a means to determine susceptibility to a phenotype resulting from an allele or marker in linkage disequilibrium with such polymorphic forms.

DESCRIPTION

I. Novel Polymorphisms of the Invention

The novel polymorphisms of the invention are listed in Table 1. The first column of the Table lists the names assigned to the fragments in which the polymorphisms occur. The fragments are all human genomic fragments. SGC, TIGR and WI respectively stand for Stanford Genome Center, The Institute for Genome Research and the Whitehead Institute.

The sequence of one allelic form of each of the fragments (arbitrarily referred to as the prototypical or reference form) has been previously published. These sequences are listed at http://www-genome.wi.mit.edu/ (all STS's (sequence tag sites)); http://shgc.stanford.edu (Stanford STS's); and http://ww.tigr.org/ (TIGR STS's). The Web sites also list primers for amplification of the fragments, and the genomic location of fragments. Some fragments are expressed sequence tags, and some are random genomic fragments. All information in the websites concerning the fragments listed in Table 1 is incorporated by reference in its entirety for all purposes.

The second column lists the position in the fragment in which a polymorphic site has been found. Positions are numbered consecutively with the first base of the fragment sequence as listed in one of the above databases being assigned the number one. The third column lists the base occupying the polymorphic site in the sequence in the data base. This base is arbitrarily designated the reference or prototypical form but is not necessarily the most frequently occurring form. The fifth column in the table lists the alternative base(s) at the polymorphic site. The eighth column of the Table lists about 15 bases of sequence on either side of the polymorphic site in each fragment. The indicated sequences can be either DNA or RNA. In the latter, the T's shown in the Table are replaced by U's. The base occupying the polymorphic site is indicated in EUPAC-IUB ambiguity code. The fourth and sixth columns of the table show the frequency with which reference and alternative alleles occur at a polymorphic site. The seventh column in the table indicates the population frequency of heterozygotes of the polymorphic site.

Fragment	Position	"Ref Allele"	"Freq (P)"	"Alt Allele"	"Freq (Q)"	"H"	"Sequence Tag"
19201	179	T	0.25	C	0.75	0.38	GGTGCACCGAAAGGAYTGGGGGATAAAATTC
19212	46	T	0.94	A	0.06	0.12	GAGACTAGAGTGACAWGTTTCAGAACCCAAA
19222	179	C	0.94	T	0.06	0.12	AGGGACTCTGCGGAAYTTTCACACCTCTTTC
19224	112	C	0.94	T	0.06	0.12	ACAGAGGAGATAATCYCAGGATGCCTGTGAA

19235	173	A	0.81	G	0.19	0.30	GTTCACAATGGTGGARGCTTCATGTAATATG
19236	54	G	0.63	A	0.38	0.47	TGGAAGGGGGAAAAGRGATGGAGACCTGCTC
19269	85	A	0.56	T	0.44	0.49	ATTTTGGAGTGTGTCWTTGGGTAGCAATGTG
19307	196	T	0.94	C	0.06	0.12	CCTTTAAAGAGACCCYTGGAAATGGGCCATG
19348	98	G	0.56	A	0.44	0.49	GACTGTTGGTCATGGRGTGACGTCCTTCTCC
19348	103	C	0.50	T	0.50	0.50	TTGGTCATGGCGGTGYGTCCTTCTCCAGGCT
19359	39	T	0.56	C	0.44	0.49	TGAATACTTTGTTTTYCATGTTCAAAAAAAG
19415	161	A	0.56	G	0.44	0.49	CCTTAGCTGATCTCARAAGTCCACCTCATGA
19591	45	T	0.69	A	0.31	0.43	ATCACATATACCTGAWATAAGGTAACTCCAA
19591	156	C	0.38	A	0.63	0.47	GTGGGGAGCTCTTCCMCTACCACTCCCCACC
19599	230	C	0.56	G	0.44	0.49	TTAAAAGTAAAGGGCSTTCCAAGAGTAACAC
19635	98	A	0.63	T	0.38	0.47	AAAAATACAGTATTAWATCTTATTGTGTAGC
19641	46	A	0.88	G	0.13	0.22	TTGTGATAAGCACTARTATTATAGTCTCATG
18012	112	C	0.20	T	0.80	0.32	GCCACTTTTGCCCCTYGTGAAGTGTTTCCTG
18014	40	A	0.90	G	0.10	0.18	TTGAATAGCTACAGARGAATGAAAGTGCACC
18036	27	T	0.43	C	0.57	0.49	GAGTCAAGTACCAAGYAAACTTCTAGAAATA
18036	97	T	0.93	A	0.07	0.12	TTTAATTCTTTCATAWCTGACAGGTCAAGTA
18046	72	C	0.80	T	0.20	0.32	TTTCAGGCCAATGTGYTGTTGGGTCTGAGAT
18052	50	T	0.40	C	0.60	0.48	CTTTCATGTACGAATYTGGTTACACATCTTA
18052	67	A	0.40	G	0.60	0.48	TGGTTACACATCTTARACAGCAGAGCTGCCT
18054	46	G	0.13	A	0.87	0.23	GAGTGGGGGAGTAAARTGGAAGCAGGGTGAC
18063	105	G	0.77	A	0.23	0.36	TAAACTTAAAATTTGRTCCTTTAACAATATA
18064	54	G	0.87	A	0.13	0.23	TAAGCTGTATTTCAGRGAATGTCACAATCAT
18078	86	A	0.97	T	0.03	0.06	TTTTTTTCAGCATCAWGTCCACTAGCCAAGT
18080	41	T	0.47	C	0.53	0.50	ATCAAACTAGTCTCTYTTGTAATTAAAATCT
18080	65	G	0.53	A	0.47	0.50	TAAAATCTACTATGCRTGTTTGACTTTTATC
18080	80	C	0.73	T	0.27	0.39	CGTGTTTGACTTTTAYTCTTATGTAAATTGA
18086	63	G	0.10	A	0.90	0.18	CAGAAAGCATACTTCRTGGCTTTGTTACACG
18091	90	T	0.97	C	0.03	0.06	CTCTAGAAGTTTGACYGGGCCTTTTTTATAC
18115	70	C	0.87	T	0.13	0.23	CCTTTGGTATTCCCTYCTTTGGTATGAAAGA
18115	71	C	0.87	T	0.13	0.23	CTTTGGTATTCCCTTYTTTGGTATGAAAGAC
18119	38	T	0.83	C	0.17	0.28	GTGGTATTACAGAGGYTTGTAAAATGGATTG
18136	78	A	0.97	G	0.03	0.06	TCTTTAGGTAATTTGRTAAGAACAATAAAAG
18142	66	T	0.97	G	0.03	0.06	AAAATAATCTATATAKCCCAATAAACTCACA
18169	115	A	0.70	G	0.30	0.42	ATCTTTCCGGAAGCTRTGGAGCACAAGCAGA
18175	27	A	0.20	G	0.80	0.32	ACGCTGCCCTTTTTARTAGAACATTATCAAA
18178	68	T	0.83	C	0.17	0.28	AGGTTAGTCTGGGGGYCGGCGGGATGGACAC
18181	100	A	0.60	C	0.40	0.48	ACACTCCCTTCAGATMCAAAAGCTTAACAAA
18190	26	G	0.90	A	0.10	0.18	CGACACAGCGGACACRTCATAAGTGGAACAA
18190	62	G	0.67	A	0.33	0.44	TGAAGCTAATCATGGRGCAAGCTCCCTGGAG
18215	78	G	0.75	A	0.25	0.38	CAGAGTTCCTGCCCTRGTGTGCGGGGGGAGA
18232	60	T	0.91	A	0.09	0.17	TTGTGATACACTTAAWGAACCCCTGAAAACC
18243	36	T	0.94	C	0.06	0.12	CAGCAGCAGAATGCAYTTTGCAGAAACACAC
18244	35	G	0.59	T	0.41	0.48	TAAGCCAGCATGGGGKGGGGAGGTGATTATG
18245	115	G	0.97	A	0.03	0.06	GGACAGAGAAACATGRCTGGGGAGTAGGCTC
18247	19	G	0.09	A	0.91	0.17	CACACCACAAACGCARGTTAGTGAGCTGCTA
18261	26	G	0.78	A	0.22	0.34	GATTGCTTTATTTAARTGAAAAGCGTGATAG
18266	97	C	0.16	T	0.84	0.26	ATGGACTATCTTCAAYTGCACAAATGATGCA
18266	119	C	0.75	T	0.25	0.38	AAATGATGCATGAATYACATTTGAGACCCGC
18266	124	T	0.16	C	0.84	0.26	ATGCATGAATCCACAYTGAGACCCGCAACTC
18268	88	C	0.75	T	0.25	0.38	TACTTCCCCCATAGAYCCTGACAATGTGCTG

18299	48	C	0.56	T	0.44	0.49	TGTCTAAGATCATTAYTTGGTTTGCCAATTT
18299	52	G	0.75	A	0.25	0.38	TAAGATCATTAACTTRTTTGCCAATTTTTTT
18299	67	T	0.56	G	0.44	0.49	GGTTTGCCAATTTTTKATCTATTTGGGTCTG
18299	77	G	0.78	A	0.22	0.34	TTTTTTTATCTATTTRGTCTGAGAATTCCAC
18299	101	A	0.38	G	0.63	0.47	AATTCCACAATTTTGRGAATTCTTTTGCCAA
18299	107	C	0.78	A	0.22	0.34	ACAATTTTGAAGAATMTTTTGCCAATTATTG
18307	76	G	0.94	A	0.06	0.12	AACTCAGTTTCCGCTRTGCTATGTAAAGCAT
18324	72	C	0.97	T	0.03	0.06	GGGGTACTGATTTATYTAGATCCAAATAAAG
18327	104	G	0.41	A	0.59	0.48	TTTCGTTAGGCTAGTRGCTGAGCCATTGTAT
18330	49	G	0.47	A	0.53	0.50	GAAATCAGGGATAAGRCTGAGGAACAAGAGG
18330	66	A	0.50	G	0.50	0.50	CTGAGGAACAAGAGGRTATGTAGGCAGTGAG
18350	48	T	0.97	C	0.03	0.06	AAAGAATGTTTTCAGYTAAATCTATGAAAAG
18357	89	C	0.66	G	0.34	0.45	CAGCCCTTAGCATCASTCATCTTCAGTCTTT
18369	58	G	0.84	A	0.16	0.26	AATCTGTCACACAATRAAATGGATAAGGCCT
18387	57	A	0.66	G	0.34	0.45	TTTGGTGACCCCATARTTTGTGGTCACATGC
18387	84	A	0.94	C	0.06	0.12	CATGCTTTAGCCATAMCATGGTAACATTGAC
18395	77	G	0.41	C	0.59	0.48	AAATTGATTATTCAASTGTGCATTGGTTTAT
18396	21	C	0.91	A	0.09	0.17	TGGTATTCTCTCATCMTTCCTTTTCGCTCTT
18398	62	G	0.84	T	0.16	0.26	AAACAACTCAAGGGTKGATAACATTGCCAGT
18399	28	A	0.16	T	0.84	0.26	CTGTATCCAGTGGCAWTTTTGGCTGCTGGTT
18399	99	C	0.34	T	0.66	0.45	ACCTCAAGGGACACCYCCACCCGACACTGTT
18409	20	C	0.44	A	0.56	0.49	TGGGAAAGAGGAAATMTTTTTCTTACTAGAG
18420	38	C	0.09	T	0.91	0.17	GGGAAAATGGGAAGAYAGAGTGAAATTAAAG
18420	108	T	0.56	C	0.44	0.49	CTCAAAAAAAATCAAYGCTTATAGCAATGCT
18425	81	A	0.06	C	0.94	0.12	GTCCTAGACAGATTCMTGCACACAACAACAG
18425	101	T	0.84	C	0.16	0.26	ACACAACAACAGGAGYGGGGGTCACACGGGC
18442	62	C	0.78	T	0.22	0.34	CAAAATAAGTTTCTGYTTGGCTGATCTGGGT
18452	38	G	0.97	A	0.03	0.06	GGAGATCGGCTAAAARAAAGCATAGTTATTA
18457	120	T	0.97	C	0.03	0.06	CACATTGGGGCCACAYAAATAGGCTAAAAGG
18462	39	A	0.70	G	0.30	0.42	ACAATGGCAGAGGTGRTAGAAACCATCTCAA
18489	102	A	0.93	C	0.07	0.12	TGCAAGGATTCAAACMGGTTATGGCAATAGA
18491	109	G	0.83	A	0.17	0.28	TAAATCCCAGAATGARGGATTACAAGAAAAT
18520	75	G	0.90	A	0.10	0.18	TTGTTCTTTTACTACRCCGGAGTGGTAAATA
18533	91	T	0.80	C	0.20	0.32	TCATTTTTCATCCTAYTTACTGAAGCCATTT
18535	107	G	0.93	A	0.07	0.12	CTTCACGGGAGAGCTRTTGTTTAAAGCAGTG
18562	29	G	0.93	A	0.07	0.12	TAATGATGGAAATATRACAAATATTCAACAT
18563	94	A	0.93	G	0.07	0.12	GGTTCACTAATGTGARGACATGGTGTGGCTC
18582	69	T	0.97	A	0.03	0.06	TTTTTCATTGTGAGAWGTGCCATAATTTATT
18612	37	A	0.73	G	0.27	0.39	TCAAGTTTGGAAATGRTATTTGCAAGCAGCA
18618	51	A	0.97	C	0.03	0.06	AGGCCGAGCTAAGAAMCGCTCAGCTTCGTTA
18619	44	G	0.97	A	0.03	0.06	ACTAACAAGCTTCTGRACAGGAGGTAACATT
18640	121	T	0.50	C	0.50	0.50	GTGGGGGGGTGCAGAYGTGTCCTCTTCAGTG
18658	52	T	0.97	C	0.03	0.06	CTGCAACTTCTGCTTYCTCCTTTGCCTCTGC
18668	76	C	0.13	T	0.87	0.23	AAAAACTAGGCAAAAYAGCAAAAAGTGCAGT
18673	29	A	0.50	G	0.50	0.50	TGTTTTAATTGCAAARACTTAATTTACAGCA
18680	75	T	0.67	C	0.33	0.44	ACTCTAGCATCTGGAYGCTCCGTTGTATATT
18683	22	C	0.87	T	0.13	0.23	GTTCAGGACTGGACTYGGTCCCTTTATTGAG
18694	41	A	0.56	T	0.44	0.49	CAGCCAGCTCTGACTWCTCTCTGTTTCTGTC
18704	99	A	0.63	C	0.38	0.47	GTTCTCCGAGGGGTAMCCAGCAGGGCCTTCA

18715	76	G	0.94	A	0.06	0.12	GAGCTTTGTACATGGRCTGGGAGACAAGGGA
18723	71	T	0.50	C	0.50	0.50	AGATTTTTGAAAGTGYAACAGGTACATAGGT
18723	94	G	0.69	A	0.31	0.43	TACATAGGTAACCAARTATATAGCTTATTTG
18723	96	A	0.63	G	0.38	0.47	CATAGGTAACCAAAGRTATAGCTTATTTGGT
18740	96	C	0.56	G	0.44	0.49	TTTACCATCATGTATSAGTAGTGGATAATTC
18740	104	G	0.50	T	0.50	0.50	CATGTATCCAGTAGTKATAATTCATTTTGAT
18741	23	T	0.88	G	0.13	0.22	GTCAAGGCTTTGGACKCTCTTCAGTCATCAG
18741	38	G	0.75	C	0.25	0.38	ATCTCTTCAGTCATCSACAGAGTATCTCTGC
18741	64	G	0.88	A	0.13	0.22	CTCTGCTCTAGACCTRCTGGAGTTCAAGCTT
18742	51	C	0.94	T	0.06	0.12	CACTTTTGCCAATGTYATCGGGTTTGGTTTT
18746	114	G	0.94	A	0.06	0.12	TTTTGTAAATATTCTRTCCACATTCTACTTC
18763	38	A	0.50	G	0.50	0.50	GTACAATGGTGTGGGRTGACGATGATGTGAA
18763	53	A	0.88	G	0.13	0.22	AATGACGATGATGTGRTATTTAGAATGTACC
18768	120	C	0.63	T	0.38	0.47	CATGTGCACCCTTGGYTTCGCTCCATCGCCT
18771	57	A	0.81	G	0.19	0.30	CTCGGAGGATGCCTARAGATGTTGGGAACAG
18771	75	G	0.88	A	0.13	0.22	GATGTTGGGAACAGARAAATAAACTGAGTTT
18790	49	A	0.56	T	0.44	0.49	GTCAACCAGGACAGAWGCATGGACAAGGGAT
18820	70	T	0.56	C	0.44	0.49	ATGAAATTCTGAGGCYTGATTTAAATCTTTC
18821	69	C	0.44	T	0.56	0.49	CCTCTCTCGGAGGCCYAGAGGCTGGGGGTAG
18821	76	T	0.38	C	0.63	0.47	CGGAGGCCACAGAGGYGGGGGTAGCCATTGT
18846	49	G	0.94	A	0.06	0.12	AGAGCAGGAGGTGCCRAAAGCTGGGAGCGTG
18851	90	T	0.88	A	0.13	0.22	TTTCCTTATTGTATTWGTAATATAGGATCCT
18882	94	C	0.81	T	0.19	0.30	ACAACATCCTCTGCCYACACAACAAAACGTA
18908	70	G	0.25	C	0.75	0.38	AAAAGGGTCAGTATGSTTAGGGAAAACATTC
18910	112	T	0.63	C	0.38	0.47	ATCACTGTGCTGCTTYGGCTCATGGCAGAGC
18919	26	C	0.50	T	0.50	0.50	CCACAGGGATTCCGGYGCCAGACCCCATTTT
18922	74	G	0.88	A	0.13	0.22	TCACCTGGACTTAAGRTCTGGCTCTAATTCA
18932	177	C	0.69	T	0.31	0.43	ATATCTTTGAGTTCAYCTTTAGTACGTGTGG
18944	147	A	0.13	G	0.88	0.22	CCCAAATGGCTAGAARTGTTTAATTAAATTT
18952	232	G	0.38	A	0.63	0.47	TTGGGAAAAGGTGTARACAGTAGCCCCATCA
18959	123	G	0.56	A	0.44	0.49	TCGAGAAAGAGGCACRGGAAGCCGTCCTGGC
18972	112	A	0.56	G	0.44	0.49	TGGGCTGGGGAAGCARTGCTTGCTGGCCATG
18984	208	A	0.94	C	0.06	0.12	GTGATGCATTTATCTMATAAAATGCTAAATG
18985	105	C	0.13	T	0.88	0.22	TACAGAGGTAGCACAYTGATTCCAACACAAA
18987	35	G	0.19	A	0.81	0.30	CCTGCCCAGCAGCCTRGTGGCCAAGCCCAGA
19016	161	C	0.75	T	0.25	0.38	TTAGATACATAGCCGYTGTATACAGAGGTTC
19016	184	C	0.75	A	0.25	0.38	CAGAGGTTCATCTCAMCTCAACACTATTGAC
19021	20	C	0.44	G	0.56	0.49	CTCTGCTGCTGTCCASACTGTCCTTTTGAAC
19034	45	T	0.69	C	0.31	0.43	GATGAGGATAGGGGAYACTTCTATTACATTA
19037	47	C	0.94	A	0.06	0.12	TCTGGTCCTAGCCACMCCTGTATGACCGCGC
19037	155	A	0.75	G	0.25	0.38	TCCCCTTACGAACACRAAACCCAGCCCACAT
19041	198	T	0.50	C	0.50	0.50	CCTTCTCAATACAGCYGCCCTTGCAGTCCCT
19042	193	A	0.81	C	0.19	0.30	TAAATAACTCTAACCMGGCTGTGTTTAGATT
19057	175	G	0.50	A	0.50	0.50	CAGATCCCCACAGCTRTCTCTTCATCTTGGT
19064	66	T	0.25	C	0.75	0.38	TGCTGGGCTGTGTTCYCGGGCTCTTCTGGAC
19066	72	C	0.56	T	0.44	0.49	CAGTGAGCCACAAGCYTTAAAACCCATGAAC
19066	87	C	0.44	T	0.56	0.49	ACTTAAAACCCATGAYCTTCAGCTGATCGTC
19066	100	G	0.94	A	0.06	0.12	GAACCTTCAGCTGATRTCCTTAGCCAGTCCA
19066	147	G	0.81	C	0.19	0.30	TGGCATATGTTCTTGSTTGGTCACCCTGTAG

19066	148	T	0.75	C	0.25	0.38	GGCATATGTTCTTGCYTGGTCACCCTGTAGC
19066	184	C	0.38	T	0.63	0.47	TTACTTCTCCATATTYGGATGCTCAATTACA
19066	239	A	0.38	G	0.63	0.47	CTTAAACGCCTTCACRGTTTCTTTTTATCGT
19067	57	C	0.88	G	0.13	0.22	GGCTGCTGCAGCCTCSCTGGCTGTGCACATT
19067	151	T	0.56	C	0.44	0.49	CTTGGGCTCTAGGTCYGGAGAATGTTGTGAG
19067	153	G	0.50	C	0.50	0.50	TGGGCTCTAGGTCCTSAGAATGTTGTGAGGG
19067	202	T	0.50	G	0.50	0.50	AGTGTTCATAAAGAAKACATAGTATTCTTCT
19076	40	G	0.69	A	0.31	0.43	AAAAAGCAGTTTTAARGTATTCAAAATACCT
19087	37	A	0.94	G	0.06	0.12	AGCTAAGCTCAAATGRTATTTAACTTCTAGT
19092	232	A	0.69	C	0.31	0.43	AAAGATCATAATTTTMATGATTAGCCGTGTA
19102	25	C	0.44	G	0.56	0.49	GTCACGCTGAGGAGASCTTCACTCAGGAGTT
19106	247	T	0.94	C	0.06	0.12	GAACCTCCTATTTTAYTGGAATTCTGGATCT
19112	212	G	0.88	A	0.13	0.22	TTGAGGGTGACAAGGRTCTCTTCAAACAGTT
19117	134	A	0.38	G	0.63	0.47	ACATAATTGCATGAARTAGCTATTTTTTTCC
19134	162	T	0.25	C	0.75	0.38	AGCCAGGGCTAGAGGYGCACGGTGGCTAGAG
19134	263	C	0.94	T	0.06	0.12	GGAAAGGGTTGATGCYATCATTTATTGAGGG
19135	20	G	0.75	A	0.25	0.38	TACCCTGCTTTGCCTRAAAGTGTCATCAATT
19139	66	C	0.88	T	0.13	0.22	TTTACACGAGGGTAGYGGCAGATGCCTGACA
19139	110	C	0.63	A	0.38	0.47	GCAGACAACACACTAMATTTTCACGGGTGTG
19144	222	G	0.38	C	0.63	0.47	GGCTCTGCTGGAGCGSTGGGAACCAAACACC
19179	170	G	0.19	A	0.81	0.30	ATAAACATATCAACCRTAGCATTAACCCATT
19183	210	G	0.50	C	0.50	0.50	GCTCTGCCCCTTGGASTGCATTTGACCTGCT
19642	52	C	0.38	A	0.63	0.47	GACACATTATCCCCCMGGGTAAACCAGGACT
19673	35	G	0.69	A	0.31	0.43	GATGAAGAACATGATRTCACTAGTAGGTAAC
19673	180	C	0.94	T	0.06	0.12	TGTGAAAACATTTTTYCTTGGACCAGCTGAA
19724	35	A	0.25	G	0.75	0.38	ATTGTAATTTGGGTARCTGAGTCACGGTGGC
19765	57	T	0.94	C	0.06	0.12	GTATACCTTGTCCTCYATGTATCTTGTCCCT
19766	31	G	0.81	A	0.19	0.30	GTACATTGGAGAAGCRTGCAGCAGCATCCTT
19766	93	A	0.94	G	0.06	0.12	ATAGGAGCCAAAAGTRGACAAACAGAAGAAG
19856	63	C	0.63	T	0.38	0.47	TCCCCCTCCTGGAGAYGCTGCGTTCCCCAGC
19909	29	T	0.94	C	0.06	0.12	CTGAATATTCTCTTTYTTAAAAATATAATTT
19911	116	A	0.94	G	0.06	0.12	AACAATGCAATTTTTRACACTGTTTTGAAAA
19946	122	C	0.69	T	0.31	0.43	AGACGCACAGAGAGGYTCTTCCTGACCCAGA
19956	141	G	0.94	A	0.06	0.12	GTCTGGACCTCAATGRCTCTCGGAGAAGCAG
19970	126	T	0.50	C	0.50	0.50	CCTGCCAGTTCCTCAYGCGGGGACCAGCAAA
19970	167	G	0.94	A	0.06	0.12	ACTGGGTTGGTCAAARTAGTCACCTTGGCCT
19984	47	A	0.19	G	0.81	0.30	CACTGACAGGTAAATRTATAACATTAGAAAA
20014	214	T	0.81	C	0.19	0.30	AGTCACCAAGCATACYTCCTGGCTCCCCAAG
20096	21	T	0.81	C	0.19	0.30	TGGGGGCATTTATTTYGATAGAGACTGGCAC
20103	168	C	0.56	T	0.44	0.49	AGCTGGGTCCTCCCCYTTCATTCTGCTCAAA
20113	60	T	0.75	C	0.25	0.38	AAGACCTGAAATACTYGGAAACAGTAAAAGC
20122	135	T	0.88	C	0.13	0.22	CATTCAAGTTTGACAYTGAAAAACCAACTGG
20146	31	T	0.88	C	0.13	0.22	TCATTGAGCAGTTAGYCATTTGAGATAAAGT
20218	26	T	0.94	C	0.06	0.12	TGGTTTTATAAAGCTYAGGACAGAGCAGAGA
20295	154	T	0.25	G	0.75	0.38	CCAGTCTATTGCCAGKCCAGAGAAAGCGCGG
20310	125	G	0.38	A	0.63	0.47	CTCTCTAGAGGCTCCRTCAGAACTGGACCCT

20907	241	A	0.63	C	0.38	0.47	CTAAAAAACATTTTTMAATTATCTAAACAAA
20964	87	G	0.44	A	0.56	0.49	GGTAGTCCACAGAATRGACACAAGAAACCTC
20993	139	A	0.75	G	0.25	0.38	AAAACCCTGGGCTTCRTAACAAGTGAGTATA
21006	106	A	0.69	G	0.31	0.43	ACACATGTGCACACARAGAGGCAAGTACAAA
21016	207	G	0.94	A	0.06	0.12	GTGCGCTGTGGGTCCRTTGGCGTGGTGATGT
21028	121	A	0.75	C	0.25	0.38	TTGAGCAATCTAGGGMTATGTGACAGGGGTT
21028	139	A	0.75	G	0.25	0.38	ATGTGACAGGGGTTTRTGCACTGGTACAGAA
21031	31	C	0.75	T	0.25	0.38	GACCTCTGACATGTGYCTCTGGTCCCCATTT
21034	148	T	0.88	C	0.13	0.22	TGGGAGATTGGATAGYGCCTAACCTATCTCA
21054	23	G	0.13	T	0.88	0.22	CTGCATGGGTACAAAKTCCAATTCATACTTA
21059	63	C	0.56	T	0.44	0.49	TTCCCACTGAGCCTGYTGAACTACAGCTGCC
21059	181	T	0.50	C	0.50	0.50	AGTCATTTCTCTATTYATTGTAGCCAGGGCA
21079	50	G	0.94	A	0.06	0.12	ATGCATGCAACTGTGRCGCAAAATCAAGTTG
21079	166	G	0.94	A	0.06	0.12	AATATCTGCTAGTGGRAATTTACAACCCACT
21122	42	C	0.75	T	0.25	0.38	AAGCTAAAGTTATTCYTTAACAGGAACTCTG
21139	165	T	0.44	C	0.56	0.49	TAAAGGAACTAATACYGTACAGCACTTCAGC
21149	167	G	0.13	A	0.88	0.22	TGGAAAGCTTTTACARTGCTTCAGAATGCGG
21155	36	A	0.75	G	0.25	0.38	TTGATGGAAAATTGGRTCTGTGTAGAATGAT
21186	95	G	0.25	A	0.75	0.38	CTGAGGTGGGGCTTARAATTAGTATTTCGAA
21190	39	T	0.56	C	0.44	0.49	TGTTTGTATAAACTAYGTGGGGTAAGCCCTT
21202	61	T	0.94	C	0.06	0.12	GTATAAGCTAAATATYTGATCTGTTTTATGA
21202	156	A	0.94	C	0.06	0.12	GGAGAGAGTTGACCAMGTCTACATGCATAGA
21235	43	T	0.06	C	0.94	0.12	GGGCAGCAGGGCAGTYCTCGGGCCGATGTTC
21242	115	G	0.44	A	0.56	0.49	GGGAGGGGCAGAGAARCACTAGCTTGGGGGT
21254	53	A	0.88	G	0.13	0.22	AACTATTCCACAGGARCAAGGAGAAGCTGTT
21376	188	A	0.25	G	0.75	0.38	TTAGGGATGAGTTCTRGAAGTGATTCTGAAC
21399	75	C	0.75	T	0.25	0.38	TCTAAAGTTTCAGTTYTTCACCAGTAAAGGA
21444	39	A	0.31	G	0.69	0.43	AAGAAATACTCTCAARAGTTCTTTTTTTATG
21485	82	C	0.69	T	0.31	0.43	GCAATTTTCATGCAGYTGTGCACACAGTACA
21504	147	C	0.81	T	0.19	0.30	TCCCAGATGCAACAAYGCGGTTCTGGCTTCT
21512	54	C	0.94	G	0.06	0.12	ACAAAAATATTTCTGSTAGAGAGGGAAAGAG
21513	192	G	0.31	A	0.69	0.43	TAAAGAGGCAGTGTARAGTAGTATTCTCTAC
21524	35	A	0.94	C	0.06	0.12	AATGAAAAGGTGTAAMGCCTGATGTACGACC
21524	97	C	0.81	T	0.19	0.30	GCATTCCTGTCTACCYGATGATGCTTCTCTC
21552	66	G	0.69	A	0.31	0.43	TACAGATACACAATGRTAATAATTACTTCAG
21552	166	C	0.88	A	0.13	0.22	ATTATTTAAAAATGTMAATTAAATTTATTAT
21561	55	T	0.63	G	0.38	0.47	CTATACCTTCGAAACKCCTCTTAACCTCTCC
21627	106	A	0.50	G	0.50	0.50	TCAACTTGAGTACCTRTTATGGATATTTATG
21627	153	A	0.94	G	0.06	0.12	GTAAGGGCATTGCAARTCCAAAGTCATCTAA
21636	71	A	0.81	G	0.19	0.30	TACCAGCTTTTTTAARTAGCAATATCTATTA
21660	120	C	0.94	T	0.06	0.12	CAAGCCTAACCTGGCYTGTCTTTTTCAGGCT
21661	117	G	0.94	C	0.06	0.12	ATTTTAAAATAAAATSTTTAGTCACAGTCAC
21703	134	A	0.31	G	0.69	0.43	CATTGGAGCCTACACRCTTGTGCTTTTCTCA
21703	197	A	0.31	G	0.69	0.43	GGAGTGAGTCTGGGARGTGGGCAGAGCACAG
21723	82	G	0.94	A	0.06	0.12	ATGGACTTTAAAGCTRACATAAAATTAGTAG
21723	125	A	0.25	G	0.75	0.38	TTAGTCATATTCCCCRCAACAGCATGATAAA
21763	135	T	0.38	C	0.63	0.47	GACTGTTCTCAGTCAYGCTCTCCCACAGCTG
21763	154	A	0.38	G	0.63	0.47	TCTCCCACAGCTGATRCAGACATTGCCTGTG
21778	155	T	0.56	C	0.44	0.49	TGGGCTTCTGAGGTCYGGTAGAAGGAGGGCA
21863	47	C	0.69	T	0.31	0.43	GCCCTGGCCCTGCCCYAGCTGCATGCCACCC
21909	153	A	0.25	T	0.75	0.38	TCTTAACATACCAAAWAGTGGAATCAATAGA
21930	146	G	0.56	C	0.44	0.49	TCCCCATTTTGAGTCSCATAGTCCATTATAT
21956	26	T	0.63	G	0.38	0.47	TCTCTTTCAAGTGAAKTTCCTTTCGTTCCTG
21961	73	G	0.13	A	0.88	0.22	TTACTTTTATTTTTCRTAAGTTATTGGGGTA
21961	200	T	0.94	G	0.06	0.12	TTTTATCCCTCGCCCKCTCCCACTTTTCCCC
21965	112	A	0.25	G	0.75	0.38	GACCTCCCCCACAGCRCCCCCACAGGGTTCT
21966	148	G	0.69	A	0.31	0.43	AGGGGATTGCAATGGRAACAGGATAAAAAGG
21980	25	T	0.63	C	0.38	0.47	ACACATTCATCAAGGYAGATTAATTAATGTC
21981	61	T	0.31	A	0.69	0.43	TCTTGAAGAAAAAAAWGTCTCCCTTATGGGT
22012	57	T	0.56	C	0.44	0.49	GCCTACATCTGGAATYCATTACATCAACGTT
22020	27	C	0.75	G	0.25	0.38	TGCAGTGCGGATGAASTTATCATGATGCTAA

22082	67	C	0.19	T	0.81	0.30	AGTTATTGGTTGTGTYGTTTTCCTTTTTGCA
22082	179	G	0.88	A	0.13	0.22	GCCGAAGGACGTATTRCTGAACTGGGACGAG
22091	205	G	0.94	A	0.06	0.12	TTACTTGAGGGCAACRAATTACGGCTTAACA
22132	99	T	0.81	G	0.19	0.30	GCCTTTTACTATCCTKCCCCATTTCTTCTAA
18017	87	C	0.25	A	0.75	0.38	GGCAACCCCNGGAACMACTGCTGGATAAATC
22202	128	A	0.94	G	0.06	0.12	TGAAATCTGAATTTCRTTAATACTCTGGTGC
22283	109	T	0.94	C	0.06	0.12	CTGCAGGCTCTGGTTYTTCATTTGCAAAATA
22292	53	A	0.94	G	0.06	0.12	ATTGCTCAGTACCAGRGTTTGAGTACGGTCG
22387	186	C	0.81	T	0.19	0.30	AAGGCAGGATTGTGGYCCTTGTGTTTTCTGA
22405	90	A	0.88	C	0.13	0.22	ATTGGCTGTAAAGTCMGATCAGGTGCTCTCC
22440	64	A	0.94	C	0.06	0.12	TTAAGCCACTTGGGTMTCCATTCCAGCTCTG
22457	112	G	0.75	A	0.25	0.38	AGGCATGAAGGATACRCAGTTAATTAACTAA
22585	56	A	0.63	G	0.38	0.47	TGACAAGTGAACAATRCAGAAGCAGCAGTGA
22631	52	T	0.81	C	0.19	0.30	CTGGCTTCAGTTCTGYAGCACCATTTTCAAG
22652	32	G	0.50	T	0.50	0.50	GCCACTTTGGAGAAAKAAGAGAATGCTATTA
22663	38	C	0.81	T	0.19	0.30	CTTCTCACTGCACTGYGAGGTGAGCCGGCGC
22663	55	C	0.56	T	0.44	0.49	GAGGTGAGCCGGCGCYGCTAATCTTATTCCC
22663	139	G	0.81	A	0.19	0.30	TGGTGCACTTACAGGRGAAGAGCTTCCTCAT
22714	212	C	0.63	A	0.38	0.47	GAGCTTACCAACCCCMTGAGTAGGGGCCAAA
22724	117	A	0.56	G	0.44	0.49	AAAGCTTGCTAAGGGRGTTATTCTATTTTTG
22750	48	G	0.88	A	0.13	0.22	AGCTGAGGCAGCTAARGGCTCATACAAAGGT
22775	60	A	0.69	G	0.31	0.43	TTCCATTGTTTACATRTAGTAGGAAAGGGAA
22808	143	C	0.50	T	0.50	0.50	ACCAGGAGGATGAAGYAGCAAACTGATTAAG
18148	101	A	0.13	G	0.88	0.22	CGATTCTGAATATCCRTGGCGGCATATGCAA
18254	64	T	0.56	C	0.44	0.49	AGAGCAGTTAAATCAYGCCAAAATTCCCTCT
18265	117	C	0.88	A	0.13	0.22	AGCCATGAACTGGCTMGTTTTCAACCTTTCC
18295	40	C	0.94	T	0.06	0.12	TGTGGAGAACAAACAYTTGGGAAGTAAAGGT
18459	64	T	0.31	C	0.69	0.43	GGGTGGGAGACACAAYGAGTAATTAACAACA
18501	121	C	0.88	T	0.13	0.22	GCAGGACAGAGGGGCYGGACAGCAGCGCATG
18548	62	G	0.56	A	0.44	0.49	AGTCCCCTCACTGGGRAAAAAAAAGCATCTN
18548	65	A	0.94	G	0.06	0.12	CCCCTCACTGGGGGARAAAAAGCATCTNTCA
18700	97	T	0.13	C	0.88	0.22	TGCTGAGAGCAGAGCYAAGATCCACAATTGC
18829	35	T	0.0000	A	1.00	0.0000	GGGGAAAAATCCTAGWAATAACTTATGTGTA
18829	58	A	0.44	G	0.56	0.49	TTATGTGTACTTCTTRTTTCATCATACAAGA
18916	35	G	0.75	C	0.25	0.38	CCAAACATCTTCAGCSCTCAGCCGGCTTCCC
18916	42	C	0.75	T	0.25	0.38	TCTTCAGCAGCTCAGYGGCTTCCCACTTCTT
19105	33	T	0.19	C	0.81	0.30	GGACAGAAAGAATATYGTGGTCCATGTGGTT
19105	211	C	0.94	T	0.06	0.12	ATCTCCCCACAACTTYTCCAGGGGCAGGATT
19576	113	A	0.81	G	0.19	0.30	AAAAAATTTAACATTRTCTAGTTCAGTGATT
19828	200	A	0.56	G	0.44	0.49	CACCACCACCCAAAARCTTTTAATTCTGGAA
19860	51	C	0.50	G	0.50	0.50	AATGTTTCCAAAGATSCTGCATCAGTATCTC
19888	98	C	0.13	T	0.88	0.22	TAGAAAGTAGCAGTGYTGGACAACGTTGTAA
19889	80	C	0.56	T	0.44	0.49	CAAGAGGAGTGAGGGYTACAGCATTTATTTC
19891	172	C	0.75	G	0.25	0.38	GCCATCTGTCTGACTSCGTCTTCCCGGGGCG
19937	185	C	0.75	T	0.25	0.38	GTGTTCCTCAGCAAGYGTCCAAACCTTCCAA
19937	186	G	0.81	A	0.19	0.30	TGTTCCTCAGCAAGTRTCCAAACCTTCCAAA
19941	71	C	0.38	G	0.63	0.47	ACAAGGTGAAAGGTASGGTCCTGGTGAGACA
20059	59	T	0.63	A	0.38	0.47	ACAGAGTGGATAACCWACATTGGCTGGAATG
20116	22	C	0.75	G	0.25	0.38	ATTTTCTGTCACCCASCTGTCCACCAGTTAT
20116	59	T	0.75	A	0.25	0.38	CCTTCAATATATGGCWTAGAACATATATAAA
20116	69	T	0.81	A	0.19	0.30	ATGGCGTTAGAACATWATAAATCTATATCAT
20155	81	C	0.75	T	0.25	0.38	CATTCCCTTGCGGGGYGCAAAACTGCTTTGA
20258	157	G	0.88	T	0.13	0.22	CCGCGGGGTGTTTCAKCGCGTTGACGCAGGT
20270	53	G	0.94	A	0.06	0.12	ACAGGAGTGGGGACGRTCAGTGTACAATACA
20270	91	T	0.31	G	0.69	0.43	TCCAGGATAAGGAGCKACACCAGGATTTATA
20317	217	G	0.38	T	0.63	0.47	AAACCATCATCAGAAKTATTAAATTAATTGC
20329	68	G	0.94	A	0.06	0.12	AGACAAGACATCAATRTCTGTTAGCAGCGAG
20442	37	T	0.63	C	0.38	0.47	AAAANGGGGGGGGGCYTAAGGTGGCACAATT
20466	133	G	0.63	A	0.38	0.47	TGAAGTGAATAAACGRTGTGAACTAATGTTT
20561	25	A	0.69	G	0.31	0.43	TTTAAGATGGCTGTTRTAAGTATAAAGCAGT
20561	94	T	0.31	C	0.69	0.43	GAAAAATCCTTACATYGGAATCAATGTCTTT

20601	125	T	0.56	C	0.44	0.49	ATTAGTCTTCTCTGTYCTTGGTGCAAGTTTG
20622	130	T	0.50	C	0.50	0.50	TATCTTAAAAGTTGAYTACTAATTTTTATGA
20768	71	C	0.94	T	0.06	0.12	CCTGCCTGCCTGCTCYGACTGATTACTTTCA
20768	190	C	0.94	T	0.06	0.12	ACACATACTGCTGGGYCAGGGACTCGTAATT
20893	179	T	0.63	C	0.38	0.47	CTGGGNAAACCTGCCYTTTCTTCTCTTTTTA
20893	207	A	0.38	G	0.63	0.47	TTTACAATGCAGTTTRACATAACATTGGTAG
20934	72	T	0.88	G	0.13	0.22	ATTTGTATTCAGAGAKTCTAAGACAAATGGT
21117	227	C	0.81	T	0.19	0.30	TCTACAGTCTGTATTYTTCTACTGAATCTTG
21187	94	A	0.19	G	0.81	0.30	CACACATAAAGACACRGGNTCTCAGTAATGC
21249	155	T	0.56	C	0.44	0.49	TCTAGGTGTATACTTYATGGAACTAGTTTAT
21314	122	A	0.63	T	0.38	0.47	CTCTGTCAAACTTTTWTTTTGTTTATAAACT
21342	59	T	0.38	C	0.63	0.47	ATNAGCAATACACTGYTGGAAATCTGCATGA
21382	125	C	0.81	G	0.19	0.30	TGGGATNTGGCTTCCSAGGTTGCAACCCCAA
21437	201	G	0.88	A	0.13	0.22	TCACCTTTACCAGGGRCAGGCATAGTGTGGC
21449	222	C	0.75	T	0.25	0.38	ACCCCTCAGCTTCCCYTGACAGAGCCAGTGT
21475	117	A	0.81	T	0.19	0.30	AAACCCCAGGCTTCTWCTTGCTTACTAAGCA
21475	181	A	0.75	G	0.25	0.38	GTCTTTGGAGAAGGCRAAAAGCCACAGCAGC
21514	100	A	0.56	G	0.44	0.49	CATTACAAAACCCCCRTCTTCAAGGAAAGGA
21514	133	C	0.13	T	0.88	0.22	CACATTACCATGGAGYACAGGACTCCAAAGG
21558	157	G	0.50	A	0.50	0.50	TGGTGGGGGGCAGTARAGCCAGGGACTCCCT
21569	198	T	0.69	C	0.31	0.43	AGAAATTTATCTCTAYAGAGACAATTCATAG
21574	235	C	0.44	T	0.56	0.49	TTACTGCCTACTTCCYGTCTGTCAGGTGGGA
21609	42	C	0.94	T	0.06	0.12	TCTCCCTTGTAACAAYGTGCAGTCCGTTCAC
21609	146	G	0.88	A	0.13	0.22	AAAGGATGTTTCAAARAGGGTCCCGGCTATG
21614	55	G	0.69	A	0.31	0.43	TTTGANTATAGCTATRTTTTAACAAACCTCA
21615	151	C	0.38	T	0.63	0.47	TTTCACTGAGTATTAYAGGACACAATCGACG
21644	151	T	0.81	A	0.19	0.30	TTTCATAAATAAGGGWTTCAATCAAGATCCA
21687	115	C	0.44	G	0.56	0.49	GGACTTTCTTCCTAASTGTTCTATGATCAGA
21695	141	A	0.88	C	0.13	0.22	CCTTTCCAAGGGAATMTACTACACTAAGCCT
21760	35	A	0.75	G	0.25	0.38	GATGCAAATGATTTGRGGTGTCTTCCTAGCT
21760	81	C	0.75	A	0.25	0.38	GGGACCTCTGACTGCMCCTCTGTCTCAGTTT
21761	138	C	0.94	G	0.06	0.12	TAAACGTGCCGTGGCSCAATACACACCAAAG
21805	45	A	0.69	T	0.31	0.43	TTTATAATCCTATATWAAAAAAAATCTATAG
21941	79	A	0.13	G	0.88	0.22	AGAGTGAGGGGCAGARGGATGAGGCTCTTCT
22129	45	T	0.50	G	0.50	0.50	AACTTTTAAGGAAAAKTTATATAACAGTCAT
22130	165	C	0.94	T	0.06	0.12	ACCCCGCGCGCTTGCYGTGTTTAATCCAGGT
22187	110	C	0.13	A	0.88	0.22	ACATTTAAAAACCAAMCAAAACAAAACAAAA
22187	178	G	0.69	A	0.31	0.43	TCTATTGGTAATGGTRAAATTTCATGAAAAT
22189	70	C	0.88	T	0.13	0.22	TGAAGTGTCCTATGAYGAGGCGAGGAATGGG
22250	89	G	0.50	A	0.50	0.50	GGAATGTGCATTCACRTAGTGGGTTATTATG
22250	132	C	0.94	T	0.06	0.12	TCCTGGCTGTGTTATYGGANCCAGGAGTGGA
22290	136	C	0.88	T	0.13	0.22	TCAGGACCTTGCTTTYTTCCAATCTCTCCTT
22374	149	T	0.94	C	0.06	0.12	TTATTCAGTAACTAAYAGGNTCCTGCATCAT
22395	127	A	0.69	G	0.31	0.43	GGGGCAACTCTTTAARAAGGAAATGTTACCA
22419	67	T	0.13	C	0.88	0.22	AGGCACAGCCCAGTGYCTGGATGGCATCAGC
22449	74	T	0.94	C	0.06	0.12	AATACAGTACTTCTTYGAAAAAATACACAAT
22512	104	T	0.94	G	0.06	0.12	GGTCCTTTGTGATCTKACCTCACCCATGTCT
22668	99	A	0.69	G	0.31	0.43	AGTTTTCTGTAATATRTTCTAGTCCATTTAG
22734	44	G	0.75	A	0.25	0.38	GGGTCTGGGAAGGCCRTCTTAGAAGACATTA
stCSF2RB	149	C	0.94	T	0.06	0.12	GGAGCCCAGAGGTTTYTGGGACTCCCAGCCA
stCSF2RB	192	G	0.94	C	0.06	0.12	CCAGCCCCAGAACCTSAGTGCTTCTTTGACG
stD22S100	88	G	0.94	A	0.06	0.12	CCTGGCAGGAAGAAGRGGATCCAGCAGTGAG
stFIBB	341	T	0.69	C	0.31	0.43	CCCACCCTTTGAGCTYACCTGCCCCACCCCA
stFIBB	412	G	0.56	C	0.44	0.49	TTGCCCTTCCCTGAASTGCCTTCTTGTGGCT
stIGLV2	61	T	0.56	C	0.44	0.49	CTCTGCTGCTCCTCAYCTCCTCACTCAGGAC
stSG10017	33	G	0.81	A	0.19	0.30	ACTCCTGGTGTCAAGRATCCTCCCACCTCGA
stSG10017	70	T	0.44	C	0.56	0.49	CAGGGTGCTGGGATTYAGGCATGAGCCCCCA
stSG10023	63	A	0.31	T	0.69	0.43	CCAATATCATTGAGGWAACAGTTTGGGCTGT
stSG10096	36	G	0.44	C	0.56	0.49	CTCCCTCCCCATGACSGGCTTCCCGCGGGCA
stSG10118	107	C	0.50	A	0.50	0.50	TGCCCCATTCCTTGCMCTCAGCCCTCAGTTC
stSG10120	89	T	0.94	C	0.06	0.12	CACGAACACTTTAATYGTTGTTGTAATCTGA

stSG10178	42	C	0.75	T	0.25	0.38	CTGGACATTAAGTCCYGGGAGGAGAAGTGAA
stSG10193	136	G	0.75	A	0.25	0.38	TATACAAACTTTTACRTTTGAAAACTGAGAT
stSG10202	143	G	0.94	T	0.06	0.12	CTGCTTCTCGCTGTCKCAAGACCACAAGGCA
stSG10209	34	C	0.56	T	0.44	0.49	CTCAGTCACCATGATYAAATAAACTAATTCT
stSG10209	75	A	0.94	G	0.06	0.12	CCCACTTTATTTTTTRCTCCAATAAATGTAA
stSG10218	29	T	0.38	C	0.63	0.47	AAATGAGAAGATTACYGTGAATATTTAAAGA
stSG10252	108	A	0.63	C	0.38	0.47	CCTTTCCCCTTGATCMAGTGAAGATATGATA
stSG10266	55	T	0.94	C	0.06	0.12	GAATTGTTCTTCTGTYGACAGTTGAAGTGGG
stSG10282	70	T	0.88	G	0.13	0.22	TGAAATCTTTACAAGKAAGCACAGTAGTACA
stSG10310	128	C	0.38	A	0.63	0.47	AAATAATTTTTTCACMTTGTCAATGCCAATG
stSG10331	107	A	0.81	T	0.19	0.30	TAGACCTCAAACACCWCACCTCCATGCATTT
stSG10331	116	T	0.94	C	0.06	0.12	AACACCAACACCTCCYGCATTTCCTCTTTGG
stSG1243	225	G	0.13	A	0.88	0.22	AAAAGAAATTCTGTTRAAAGTATTTCAGACC
stSG1345	54	T	0.50	G	0.50	0.50	TTTGAACTAGTTTGCKCTTACGCGCTTCACA
stSG1345	60	G	0.63	A	0.38	0.47	CTAGTTTGCTTCTTARCGCTTCACATTTTAG
stSG1385	117	T	0.94	G	0.06	0.12	GAGACTTGGTATTTTKTCAATCATTAAGAAG
stSG139	69	T	0.19	C	0.81	0.30	ACAGCACTTTGTGTCYGCTTTGAGCACTTGC
stSG1427	103	T	0.25	C	0.75	0.38	TTGGCTTCTGTCCTCYAGTCTCTCTCCATGT
stSG1471	50	A	0.13	G	0.88	0.22	GTCATGTGTTAGGTCRCTCCCTTGCATGAAA
stSG1483	44	T	0.06	C	0.94	0.12	TACTATTTAGTCTAAYTTTAATTCAAAGGTT
stSG1696	67	C	0.94	G	0.06	0.12	GCAAAACCAGTGTGCSAATGTGGAGGATGTC
stSG1847	49	C	0.38	A	0.63	0.47	CAACACAAATGCTACMCTAAAATGAAAGAAT
stSG1847	95	G	0.38	A	0.63	0.47	AAACAAGTGAGAGACRTTTACTTACATCAGT
stSG1897	83	A	0.56	G	0.44	0.49	AGGAGGACACAGGACRGCCCACCACCTTCTC
stSG2022	86	T	0.00	C	1.00	0.00	TTAACATTAATATACYATTCCATAATCTCAT
stSG2034	166	T	0.81	A	0.19	0.30	AAAATAGTACATGTTWGTGTAATAAAATTAA
stSG2076	104	C	0.94	G	0.06	0.12	AATATATTTTGACATSACATCACAGTGGGGC
stSG2108	49	T	0.19	C	0.81	0.30	CCAACCAAAAAATGAYGAGGGGCTCCACAGA
stSG2108	71	A	0.19	G	0.81	0.30	GCTCCACAGAGAGAGRTAAGGGGAAGACTTT
stSG2141	113	C	0.94	T	0.06	0.12	ATGGCAGCACCACTGYATGGCGATGGTGCAG
stSG2141	173	A	0.75	G	0.25	0.38	GCTTGAAGAGAGAAARAAGTTCCCTATTATT
stSG2148	50	A	0.88	G	0.13	0.22	TTTAGACCGTGATTTRAAAGAAACAATAAAT
stSG2175	68	C	0.94	T	0.06	0.12	AAATCTGTTGTGTGCYGCCGCGTGACTCAGC
stSG2189	41	C	0.69	T	0.31	0.43	CCTGATATTCACACTYCTACATTCCCTCCAG
stSG2200	49	T	0.25	C	0.75	0.38	CTGGTTCTGTATGATYTTTATATTTATGTAT
stSG2218	48	C	0.81	T	0.19	0.30	AAGAAAAAAATCCTCYTTAAAAAAAACAAAA
stSG2218	139	G	0.94	T	0.06	0.12	GCATTTTGGAATTTAKGTTTGAATAAAATAC
stSG2218	201	A	0.44	T	0.56	0.49	AAACATTCTGGTATGWTATTGTGAGTGGTGC
stSG2243	85	G	0.81	T	0.19	0.30	ATGGTCAGTAGAAAAKAGAGCATCTCCTCAG
stSG2257	65	A	0.94	C	0.06	0.12	GCTATCAGAAGGGCAMCTGTCAGGAACTCTC
stSG2306	67	A	0.13	G	0.88	0.22	TGGGAACTATTTTACRTATGCTCCCATTGGG
stSG2334	70	T	0.38	G	0.63	0.47	CGCAAAAAACAAAAAKTGCAGTGGAGGGGGC
stSG2339	63	T	0.44	C	0.56	0.49	AAGTAACTGCTGTCAYGTTCTCAGAGTCACC
stSG2465	76	C	0.13	T	0.88	0.22	CAAAATGCAGAAACCYTACAGATTAAAAGAG
stSG2549	140	T	0.69	C	0.31	0.43	GCAGCTAAAGGAATAYTACACCACCCACCCC
stSG2577	121	C	0.13	T	0.88	0.22	AACCGAACTGTGAAAYATGAACAATCCCGGC
stSG2577	123	T	0.88	G	0.13	0.22	CCGAACTGTGAAAGCKGAACAATCCCGGCCC
stSG2700	58	G	0.31	A	0.69	0.43	TGAACTGTCCGGCCCRAGTCACTCAGCGTTT
stSG2724	101	T	0.38	G	0.63	0.47	ATTGCTTGCATAATCKTTTTTTTAATCCTGG
stSG2776	65	G	0.50	A	0.50	0.50	AAAGTCTCGAATATGRTATTGGCCCTTTTGG
stSG2791	100	A	0.44	G	0.56	0.49	TAAACTAGCAATTTTRTAAATATTGGGGTCC
stSG2791	109	G	0.88	T	0.13	0.22	AATTTTAATAAATATKGGGTCCACTTAAATC
stSG2826	85	C	0.50	T	0.50	0.50	CTCCCTCCAAAACAAYGAACAAAAATAAAGA
stSG2850	88	G	0.56	A	0.44	0.49	CCCAAGGGAGACGGCRGGCTCACACATCCCA
stSG3031	71	T	0.94	C	0.06	0.12	CTGTGGTGTCAGCAAYGCCCCTTTATTTTAA
stSG3058	81	G	0.75	A	0.25	0.38	TGAAAAAAGTCAAAARTGAAGAAGCATCAAA
stSG3092	94	T	0.94	G	0.06	0.12	TAATAAATGAACGTGKGATAAACATTCTTCT
stSG3230	95	A	0.63	G	0.38	0.47	AATTGTCAGTGGAGTRGTGGGGTGCTAAGTG
stSG3245	160	G	0.81	C	0.19	0.30	CCTACCTGGGAGGTTSTGTACTTGGCTTAAG
stSG3265	42	T	0.88	C	0.13	0.22	ATTTATTTATAAGGAYGCATTGTGAATAGTT

stSG3269	24	A	0.50	G	0.50	0.50	CTGTGTCATCCTATCRTTCCCTTCCCTGAGC
stSG3269	141	C	0.81	T	0.19	0.30	CCATGCTAAAGCATGYTGTAGATCCCCAAGT
stSG3284	130	C	0.75	T	0.25	0.38	CACTCAGACTTCCCCYTCCCTAACTTTTGTT
stSG3292	99	A	0.63	T	0.38	0.47	TGACTTAAATATCTAWTACAAATCAAATAGC
stSG3323	26	C	0.94	A	0.06	0.12	ATCTTTAGCTCTCACMCCAGTGTATCCATTT
stSG3369	69	C	0.63	T	0.38	0.47	AGGACCACTCAGAGGYATAAGGGAACCCTCT
stSG3398	125	G	0.56	T	0.44	0.49	CTGTCACCTTTGTAGKCTGGGTCAAAGTCTA
stSG3416	43	A	0.06	G	0.94	0.12	AAAGGATGCAATCACRCTCACTGTAGCCTGG
stSG3424	173	T	0.44	A	0.56	0.49	TGCTGGGTAACACTGWCAAGTTGCTTAACCT
stSG3436	88	T	0.31	A	0.69	0.43	TGGCAGAGAGGGCCCWGAAATAGCTTACTCT
stSG3463	103	C	0.19	T	0.81	0.30	CAGCTCAATGGGTCAYTGGAACAAACTTGCT
stSG3470	123	A	0.81	C	0.19	0.30	TTACGATCATTTTAAMATTTTAAGAACTGAG
stSG3491	71	G	0.81	A	0.19	0.30	AAGGACGATTTGAAGRGTGGAATTACTGTGC
stSG3492	71	G	0.88	C	0.13	0.22	TAAGGCCATTCTGTGSTTATTTTTAAAACTT
stSG3523	33	C	0.63	T	0.38	0.47	TTCTTTTTGGGTTTTYGCATATATGTGTGTA
stSG3536	213	A	0.63	G	0.38	0.47	GCTTGTGCACCATTARTCCTGCTGGGTGTTC
stSG3583	112	G	0.88	A	0.13	0.22	ACATCCACACAGGCARTAACATACACAGTAC
stSG3586	60	G	0.94	C	0.06	0.12	ATCAGGTGTGGTGGTSACGCCTGTAGTCCCT
stSG3589	101	T	0.13	C	0.88	0.22	CAAAAAACCCCAATGYCCTATTTCCAAGAAT
stSG3590	70	A	0.81	T	0.19	0.30	GTTCTAAAAAAAAAAWTTTCTCTGATGTCTC
stSG3619	78	A	0.88	C	0.13	0.22	TACGCTTCTGTCATTMAACAAACTTCCAGAG
stSG3644	40	T	0.94	C	0.06	0.12	CATATTTAGGATGAGYGGATTGAGAGGCATG
stSG3646	43	A	0.63	T	0.38	0.47	TTGGCAAGAATATATWTGATAACAATAATAT
stSG3646	55	A	0.81	G	0.19	0.30	TATGATGATAACAATRTATGTCTTACTGGTG
stSG3646	70	G	0.38	A	0.63	0.47	AATATGTCTTACTGGRATATTAACTTTGATA
stSG3693	30	C	0.88	T	0.13	0.22	CATTCCCGTGGTCTCYTCTGAAAGCCGATGA
stSG3693	85	A	0.75	C	0.25	0.38	AAATATCCTACGAGGMTCGCCCTCCGAGACT
stSG3698	51	C	0.88	G	0.13	0.22	CCAATCCCCAGGGTTSTCTCTGACTTCCACC
stSG3698	145	G	0.88	A	0.13	0.22	CTAAGTCTTTATTGGRAGAATACCCACCCAC
stSG3724	107	C	0.88	T	0.13	0.22	GCTCAGTGATGTGAAYACACAGGAGTCCCTC
stSG3725	104	G	0.56	A	0.44	0.49	CAACAGCAAACAGCCRAGCAGGAATCGGCAC
stSG3751	128	G	0.56	A	0.44	0.49	GAGAGGATATGGTCCRTTGCTGACTCCATGT
stSG3787	49	T	0.44	A	0.56	0.49	AGCAGAGGATCTCTTWAAAGTTCCCTAAGAC
stSG3880	36	G	0.56	C	0.44	0.49	CCAGAGACCAGGGCTSGGCAGCTGGGGGTCC
stSG3880	115	G	0.50	C	0.50	0.50	CTGGGGAGCAGGTCTSGGCACGGAGGATGCA
stSG3895	44	A	0.88	G	0.13	0.22	GTATTGTTAGTGTTGRTTTTTTTTTCCATTA
stSG3902	104	T	0.88	C	0.13	0.22	GAACTGCTTTCTTTTYTCAGCTCAATAGCTT
stSG3935	50	G	0.88	A	0.13	0.22	AACAAGCAATTTGTCRCTAGTGTGCAGGCTC
stSG40	25	A	0.75	G	0.25	0.38	GTTGAAGAAGTGCTGRAAATATATTTAAGAT
stSG4009	32	A	0.69	G	0.31	0.43	GATGAATGGCGCGCTRTACTCTTTACGGTCT
stSG4033	123	T	0.75	C	0.25	0.38	AGCATAAAGGTACTTYTGTGAACAGGTGGGC
stSG4038	29	G	0.88	A	0.13	0.22	GTACAGCCACGCCTGRCGCAGGCCCACTCTG
stSG406	53	T	0.88	C	0.13	0.22	AGCTAAACGAACAAAYGGTTTTAGTTTTGCT
stSG4095	27	A	0.81	C	0.19	0.30	ATTAGTCAAGCAGGTMGATACTATTGTCTGC
stSG4095	55	G	0.81	T	0.19	0.30	CTGCTAGATGTATTAKATAAAAAAGTTTGCT
stSG4120	65	G	0.94	A	0.06	0.12	ACTTATGGATAATCARCTTTTTCCCCTCAGA
stSG4128	54	A	0.88	G	0.13	0.22	CTTTGTGTACATTTCRTATATTATTTTACTT
stSG4209	65	G	0.81	A	0.19	0.30	CATCCACATGGCACARCAGGGCCGGCCACTC
stSG4209	128	G	0.88	A	0.13	0.22	GAGGCCGCACTCCCTRGCAGGGGGACCACGG
stSG4254	31	G	0.56	A	0.44	0.49	CATGGAGGACCAGAGRCCACGGCCGGGACTC
stSG4301	81	T	0.38	G	0.63	0.47	ATTTAAGCAAATAAAKAGCTTCTGAGTAGTT
stSG4331	71	T	0.25	G	0.75	0.38	GTTTTATGACACAGAKTTTTCAAACAAGTTT
stSG4340	76	G	0.56	A	0.44	0.49	AAAACCACATGTTCTRTAAGTGGGAGATAAA
stSG4361	24	T	0.81	C	0.19	0.30	CATTGAGTGACAGAGYCAGTCATGCAGAACT
stSG4361	109	A	0.75	C	0.25	0.38	TAACTGCATCTTTTGMCCTTCACAACTAGAA
stSG4376	73	A	0.63	G	0.38	0.47	TCCAAGGGGAGAACARCTGGAACTGCGGCTC
stSG4381	50	T	0.94	C	0.06	0.12	AACATACGATTTTCTYTCAGTCTTGTAGTAT
stSG4410	79	A	0.69	G	0.31	0.43	ACCATTCAAACACCGRTGACAACGAACCCAG
stSG443	65	C	0.69	T	0.31	0.43	GGCAGTGAACACATCYGTATGCAATGAGAAA
stSG4430	54	A	0.94	G	0.06	0.12	AGTAGTTCTATAAGGRATTAACATAGGTAGG
stSG4448	99	G	0.94	A	0.06	0.12	CCTCTGGGGTCACTGRTGGGTTAGGCCCCCA

stSG4449	92	T	0.63	C	0.38	0.47	GACAACTTAAAACTTYTAGTGACATTGCTGT
stSG4465	60	G	0.94	A	0.06	0.12	CTGCACACTGGAAGGRAAACCTGGGAGAGAG
stSG4467	42	C	0.94	A	0.06	0.12	CTGGGACAGAGCCTCMAGATGATGTCCATGT
stSG4469	74	C	0.63	T	0.38	0.47	GCTTCTTGCCAGGCTYTTAAATTGTGCTGTA
stSG4475	21	A	0.81	C	0.19	0.30	TCATTTCCTGACCAGMTATTAAATAGTTTAT
stSG4477	32	A	0.94	G	0.06	0.12	GGGGGTGAGACAAACRATGAACCAATAATTA
stSG4531	79	C	0.94	T	0.06	0.12	GGGACAGCAGGCGTCYGCCACGTCCTGGCGT
stSG4550	85	C	0.56	G	0.44	0.49	AAGAGACAGTGGGCASGCAATTGGAGGGGAA
stSG4550	86	G	0.81	A	0.19	0.30	AGAGACAGTGGGCACRCAATTGGAGGGGAAG
stSG4551	74	C	0.75	T	0.25	0.38	CTCAATGCAATAGAAYTGACATGGGGCCAAA
stSG4590	47	A	0.94	G	0.06	0.12	AAAAGCTCTTCTGCARATGGGAGGGAGACAC
stSG4617	125	C	0.75	A	0.25	0.38	GAGATGATTCTTCTCMCCCTTCTCTCAGGGT
stSG4623	22	T	0.56	C	0.44	0.49	TATCACCCAGCGCTGYCAATGTACTAGTAGC
stSG4843	102	A	0.94	C	0.06	0.12	CTAAATTTTGAGTCAMATCAGAAAGTCTTCC
stSG4850	38	C	0.88	T	0.13	0.22	GAGGAGGAAGGGGCTYGTGCACTTGCAGGCC
stSG4879	86	A	0.38	G	0.63	0.47	CCTGGGACTGGAGCARCTTGGGTGAGCTCTA
stSG4885	104	G	0.88	A	0.13	0.22	ACGACTACGCTCTGCRGTGGGAAAGCAGAAG
stSG4896	112	C	0.75	T	0.25	0.38	GCTGGGCACCTTTTCYCAGCCACAGGCCCCT
stSG4932	22	G	0.44	A	0.56	0.49	CCGATGGTTACACAARTTGTAAATGTATTTA
stSG4950	24	A	0.88	G	0.13	0.22	CCCAGGAAAAGGTCCRTCTTAGCTTCCTCCT
stSG4957	136	G	0.75	A	0.25	0.38	AGGATTCATGAGCCCRGTGACACAGATGGGG
stSG4961	91	C	0.88	T	0.13	0.22	GATGAAAAGGAAAGTYAGAGAGGGCATTCAG
stSG4967	72	A	0.06	G	0.94	0.12	TAGGAGTGCAAGGGCRTACCCCCGGAGCTAG
stSG4997	22	T	0.81	C	0.19	0.30	AGAGTAGGAGCCCCAYTTTTAATGGTTTCCT
stSG50	125	C	0.44	G	0.56	0.49	GTTCCGGACCTAGATSTGACGAAGGTAGCAC
stSG6312	37	C	0.94	T	0.06	0.12	CTTTAGTGCAAAAACYTATGCCATGCGGGAA
stSG6345	107	G	0.63	A	0.38	0.47	GTGATGTTTTGTCCARATAGTTCAGGCAATT
stSG6362	88	G	0.94	C	0.06	0.12	ATGAGCACTGTATGTSAGAAAAGGGAAGGAG
stSG8010	62	G	0.81	T	0.19	0.30	TTTGGGTGTGCACTGKTGTCTTTCAACTGGG
stSG8022	53	G	0.25	A	0.75	0.38	GCCTGAAATGGACCARGTGGGAGTTATTTAC
stSG8032	67	G	0.31	C	0.69	0.43	TCAGAAAATTGTGTGSTGGGAGGCAGGGTAG
stSG8064	23	G	0.94	C	0.06	0.12	TCTTCCTTCTGTGCGSTTCGGGAGGCTTCAC
stSG8064	46	C	0.81	A	0.19	0.30	AGGCTTCACGTCCTCMCCGTGGTCCCTGGGT
stSG8072	59	A	0.69	G	0.31	0.43	TCTTGCTGTCTTAGGRTGGCAGAGGCAGAAG
stSG8100	40	A	0.94	G	0.06	0.12	CTTGTATCAAATTCCRAAGTGTAACTAAAGT
stSG8102	138	T	0.75	C	0.25	0.38	ATACAATGTGAAATGYTGTCATAATCATAAT
stSG8105	110	A	0.75	G	0.25	0.38	AGGCCTGAGAATATTRTTTCTAACAAGTTCC
stSG8130	36	C	0.81	G	0.19	0.30	GGAGGGAAATAAATGSTGGATGGTCGCTGCT
stSG8130	96	T	0.19	C	0.81	0.30	AAGCGGTGCCTGAGCYGTGCCTGTCTTCAGA
stSG8145	97	C	0.81	T	0.19	0.30	AGAACACAATTGTGAYACAAATCTAAGAAAT
stSG8145	124	T	0.81	A	0.19	0.30	GAAATGAATGAGATGWCTGAAATCTGATTCA
stSG8150	36	A	0.94	G	0.06	0.12	GATTTTTCAGAATAGRATAAATAATAACGGG
stSG8340	30	C	0.81	T	0.19	0.30	AGAGCTGGGCAGGATYCAACATTATGACCCT
stSG8416	65	A	0.63	G	0.38	0.47	CAGGCTGTCCTACTCRTGTGGTTTGCTAGCC
stSG8465	56	A	0.88	G	0.13	0.22	TCATGGGGCAAAAGTRCTATGGGGCCAGACT
stSG8466	111	G	0.94	A	0.06	0.12	GGTATTTGCACTACCRTGAAGCAGCACAGCA
stSG8656	44	C	0.94	T	0.06	0.12	ATGACCTTGATGCCGYGGAATTATATTCAGA
stSG8880	28	C	0.94	T	0.06	0.12	CTGTACCCCCGACGTYTCCCCTGCTCGGCAC
stSG8904	35	G	0.88	A	0.13	0.22	TCACGCTGATCCAGCRGGCACCCTGCTTAAG
stSG8917	64	G	0.75	A	0.25	0.38	GTAACTATGACTAGARAGGCAGAGGAGTGGG
stSG8944	30	C	0.44	T	0.56	0.49	TTGTAAGGATGTTTCYATAGAAATCACGGAT
stSG8944	48	T	0.69	C	0.31	0.43	TAGAAATCACGGATAYATCACCAGTCTACAG
stSG8944	59	T	0.38	C	0.63	0.47	GATAGTATCACCAGTYACAGCCACTATCTAT
stSG90	40	A	0.25	G	0.75	0.38	CGAGGAGTAGCCAGGRGGCGAGACACAAAAA
stSG90	69	G	0.25	C	0.75	0.38	AAAGGCCTGGGACAGSTCAGTACAAGTCAGG
stSG9044	67	G	0.56	A	0.44	0.49	TGACTGTAGAGGATTRATGATCCCTGAATAC
stSG9062	83	C	0.38	G	0.63	0.47	GTACAGCAGGCTCTASCATTCTCTCTCTCTT
stSG9073	88	G	0.75	A	0.25	0.38	CTGGGCATGGCCGTGRCACCCTGTGTGGCGA
stSG9075	65	C	0.94	T	0.06	0.12	GATTCTACAGCACGCYGACACTAACACATCA
stSG9354	41	C	0.25	T	0.75	0.38	CCAGGGAGACACCTCYGTGAGATGACCTGCA
stSG9535	42	C	0.75	T	0.25	0.38	CTCCCTGCCAGTCTTYCCGTCTAACCCTCAG

stSG9615	38	A	0.56	T	0.44	0.49	GGTTGGAATGTGTATWCAACTTGATGATGAA
stSG9615	156	A	0.56	G	0.44	0.49	AATAAGTGTTGTGAARGATTTTATTATAAAT
stSG9673	82	A	0.88	G	0.13	0.22	TCCCTCTAATGAAGGRAAGGGTTTTTGAACA
stSG9757	195	G	0.94	T	0.06	0.12	TTAATCATTACTATTKCAACTCCGTATTTTC
stD22S972E	20	A	0.56	G	0.44	0.49	TCCAGGAGCTGTTATRCTACCAGTTTCTGGC
stSG10082	48	G	0.88	A	0.13	0.22	ACTGGCAGGGATTTGRTATCTAAACATAGAA
stSG10082	58	A	0.75	C	0.25	0.38	ATTTGCGTATCTAAAMTAGAAAAGGTACAGT
stSG1398	73	T	0.81	A	0.19	0.30	TTGCTTTTTATAATTWAAAGCAAATAACACA
stSG1437	71	G	0.25	T	0.75	0.38	AAGTTTTGACTTTGGKTCAAGTTTTTATTAC
stSG1446	106	T	0.50	A	0.50	0.50	TCAGCGTGAGATGATWTTGATTAAACTTGCT
stSG1446	147	G	0.75	C	0.25	0.38	TAGTCAAACGTCGAASTTGCTTGAGATGGCT
stSG149	107	G	0.19	T	0.81	0.30	TTGGGGAGGAACCATKCTCCNTCTGGGCCGC
stSG1514	78	T	0.81	G	0.19	0.30	TGGGTTTCTGTGGGAKCAGCGGGGGCCTCCT
stSG9800	134	C	0.50	A	0.50	0.50	TTAGTTTGGATTAATMGACTTAAGAAAACAA
stSG9828	32	G	0.88	A	0.13	0.22	ATTATGTGTTTCAGARTTATTAAAAAGGCTA
stSG9889	128	C	0.94	A	0.06	0.12	AGGAACTGAGAAAGAMCTGCCTAAGCAGCAC
stSG9950	139	G	0.88	A	0.13	0.22	AAAATACTTGGTTAARTTGAAAGGACCTAGT
stSG9961	33	A	0.19	G	0.81	0.30	TCTATTAGATAAAATRCAGATAAAGAATCTG
stSG9961	45	T	0.63	C	0.38	0.47	AATAACAGATAAAGAYCTGGAGAAAGGCTTT
stVPREB	30	G	0.94	A	0.06	0.12	ATATTTCTCACAATCRACAAGAGCCAGGGCC
stSG1615	57	T	0.58	C	0.42	0.49	GAGACATCCAGCCCAYTCTCTGGAACAGGAA
stSG1615	79	T	0.75	C	0.25	0.38	GGAACAGGAAAGATGYCGGGGAGGGAACACA
stSG1615	88	G	0.42	A	0.58	0.49	AAGATGATCGGGGAGRAACACAGGTCAGTNT
stSG1615	119	G	0.50	A	0.50	0.50	TGGGGACAGGGGTCARGTGGACACGGGGGTG
stSG1813	41	C	0.50	T	0.50	0.50	GTGAGGGCCCAGGGTYTCCACGGAGAGGACA
stSG1828	191	G	0.50	A	0.50	0.50	TGCTGTAGCCAAATTRTTGTCATACCAGGAA
stSG2020	51	C	0.75	T	0.25	0.38	ATTAGAAAAGGACGCYCTGTTGGCTGAACAA
stSG2125	55	A	0.83	G	0.17	0.28	TTTACAAAATTTCATRGAACTGACAATGTTA
stSG2294	139	T	0.92	C	0.08	0.15	AACACTGCAAAAACCYTCAAGCATAAAAAAG
stSG2314	89	T	0.75	A	0.25	0.38	ATGTCTTTTCCCAGTWGTCATATTTTTGTCC
stSG2417	84	T	0.83	C	0.17	0.28	ACTCTCTTATGACAAYAGTGATTGANCTCTA
stSG2482	121	A	0.08	T	0.92	0.15	TGANGCAGGCTATGGWTAAAAGAAACAACAA
stSG2623	77	C	0.92	T	0.08	0.15	TTGTCTTTTTTTTTCYGGCAAACTTTCTGCT
stSG2679	39	A	0.58	G	0.42	0.49	TACATTAATTTTCTTRTGAACACAGTAGACA
stSG2773	49	C	0.83	T	0.17	0.28	ATATACACTGTTTATYTTTTTCTTTTTCACG
stSG3009	88	C	0.92	T	0.08	0.15	TTACTTTTTATGTAGYTAAGTGTGTTTATAA
stSG3094	79	C	0.75	G	0.25	0.38	CTCCCCAGAGTAAAASGTTTTCTCTGGGNCT
stSG3234	74	C	0.94	G	0.06	0.12	TCTNCAGATTGCACTSTAAGATCCTAGTTAC
stSG3248	40	A	0.38	G	0.63	0.47	ACATTCAGAATTATGRAAACAATTAGTTATA
stSG3277	43	A	0.75	G	0.25	0.38	TCATTTGCTACATGARCAGAGGCAGAGTATT
stSG3349	141	T	0.31	A	0.69	0.43	CCTTTTAAAAAATGTWGAAATTTTAAGTGGG
stSG3388	28	T	0.94	C	0.06	0.12	AGTGAATTAGGGAGTYCTTGTTGACCCCTTT
stSG3552	40	G	0.56	A	0.44	0.49	AAAACCACATGTNCTRTAAGTGGGAGATAAA
stSG3809	87	T	0.44	C	0.56	0.49	AGTTACAGCCCCCTCYCACTCCTGTATCTGC
stSG3809	122	G	0.63	T	0.38	0.47	GGGTGGTGATGTGTTKGCTCCTAGACTCTCT
stSG3809	123	G	0.88	C	0.13	0.22	GGTGGTGATGTGTTTSCTCCTAGACTCTCTC
stSG3885	36	G	0.06	C	0.94	0.12	ATTTCTGACATTCATSCCAAAGANGGCAAAG
stSG3927	84	A	0.94	C	0.06	0.12	ACAAAATAAACCGCTMGTTTTTCTGGCTCCA
stSG3927	118	T	0.0000	C	1.00	0.0000	CACGCCATATGAAGCYGCCAATGTCACTTAT
stSG4025	41	G	0.88	A	0.13	0.22	ATCAACAGCTGCTACRTTCACCCCAGAGGTG
stSG4044	22	A	0.44	G	0.56	0.49	TAATATGGGGGGTCTRAACACAGCACCCCCA
stSG4085	30	A	0.94	C	0.06	0.12	GCCCCAGTGATTCTCMTACATTTTTCACCTC
stSG4085	97	C	0.69	T	0.31	0.43	TTTTCTTGCCTGGAGYTTCATTGTTCACCCT
stSG4148	68	T	0.38	A	0.63	0.47	GATAAGCAGATCAGCWGCCAGCCAAGCTCAT
stSG4389	52	G	0.38	T	0.63	0.47	GGCAGTATTTTAAAAKATTTCTCTAATGTTT
stSG4494	71	T	0.94	C	0.06	0.12	ATTATTTCAGTCATCYTAACATGTGACTTTA
stSG4537	42	G	0.94	T	0.06	0.12	CCTCTGGCGAGCCCTKCGGCTCCACATCCTC
stSG4702	124	C	0.94	T	0.06	0.12	ACAGTTGCTGACTCCYGGCCTGTGGGCAGNC
stSG4978	102	C	0.44	G	0.56	0.49	AGAGAGGCATCACTGSGCTGCATCTGCCATG
stSG6328	117	G	0.50	C	0.50	0.50	GCTTTAACAGAAACTSAACTCTTCACGCTTG

stSG8971	95	T	0.88	C	0.13	0.22	AGCAATTTAATACAGYGAAAACAAATACAAT
A002Q12	26	T	0.25	C	0.75	0.38	TCATTATTCTCTCCTYAGATTATTAAATATT
A002Q19	32	C	0.75	G	0.25	0.38	TCCTTCCCCTCCTGCSCCACACTGCTGGCCA
A002Q20	138	T	0.88	C	0.13	0.22	GCCTGCATTTGGCTTYGTGCCTGAAAAAGAA
A002S01	83	A	0.69	C	0.31	0.43	AAATGAAGATTTAATMTTCCTAAATTTAAGT
A002T26	86	C	0.81	T	0.19	0.30	CGTCAAAAGAAAACCYCCCGGGACCCACTGT
A002V42	50	T	0.06	C	0.94	0.12	TTTCATTCAGTTATAYCTTTGGCTCAGCTAG
A002Y34	89	A	0.88	G	0.13	0.22	TAACAGAAACGCCTTRGACACTATGTTTGGG
A002Y45	85	C	0.75	A	0.25	0.38	GTGTGTCAGGATGCAMTGAAAGCCCTCGGCT
A002Y45	106	G	0.38	C	0.63	0.47	AGCCCTCGGCTCGGTSTTAGCCAATCTTCCT
A003B21	49	T	0.63	C	0.38	0.47	GACAACTTAAAACTTYTAGTGACATTGCTGT
A003B21	120	T	0.63	A	0.38	0.47	TTAAAAGAGCAAAGTWCCCCTCCCTTTCTTA
A003B29	68	G	0.88	A	0.13	0.22	TTTTGGCCATAGACARTTATTTTGATTCTAA
stSG9569	191	A	0.19	T	0.81	0.30	ATATGTATATATATAWTTTTTTTAATTCCTC
stSG9574	43	T	0.81	G	0.19	0.30	TTGGGGCAAAAGAGTKTCTTCATTATCAATC
stSG9792	105	G	0.75	T	0.25	0.38	CTGGTGCGCTGAGGCKTACACACCGGCAGAA
stSG9792	108	C	0.94	T	0.06	0.12	GTGCGCTGAGGCTGTYACACCGGCAGAACAG
stSG9915	81	T	0.94	C	0.06	0.12	CAAACCCATTTAAGTYGGAATGATTTATATG
stSG9997	99	C	0.88	G	0.13	0.22	GCCCTAATAATCCAGSATTCCTNACTCTCTT
A004A22	125	G	0.94	A	0.06	0.12	TATCTGGCGAGGAGGRCGGCATGGAGTCCAG
A004A30	135	G	0.94	C	0.06	0.12	GAATTTTAGATGCAGSATCATTTTATATATA
A004B17	146	T	0.25	C	0.75	0.38	ATTGCTGGGGCTCTAYTCCACAATTTGTTTT
A004B36	107	A	0.94	G	0.06	0.12	TTGGAGTGCACTGGCRTCCCTCAGATTTGTC
A004B39	58	G	0.94	T	0.06	0.12	CCTCCCCTCCAGACCKCTCCTTCTCCCTGCT
A004F06	71	C	0.94	T	0.06	0.12	ATAATTTATACCACAYCTGAAGAAATTATCT
A004F17	47	G	0.94	A	0.06	0.12	TATGGACTATGTACARACAATACAAGAGGCG
A004G25	85	C	0.94	T	0.06	0.12	AATCATTCTCTCTTCYTTCACATGGTGTACT
A004H43	35	C	0.75	T	0.25	0.38	GGACTTCCTAGCCTTYAGCAAGCTTAGAGGA
A004H45	26	T	0.75	G	0.25	0.38	GAAGTTGCTATAGGTKCTCTTTCTCAAAAGT
A004I05	49	G	0.06	A	0.94	0.12	TAATGCCATTTTGATRTTAACATTACGTGTC
A004I26	62	A	0.94	G	0.06	0.12	AAAAAGATTAAAAACRAAATAATATTAAAGG
A004I35	45	C	0.13	T	0.88	0.22	CCCCAGGTTAAACACYTGTAATTTACCTTGA
A004I36	173	A	0.63	C	0.38	0.47	AGACAATGTCACTTGMAACACAAGGTATGAA
A004I36	190	T	0.31	C	0.69	0.43	AACACAAGGTATGAAYATAAATAATAGTCAG
A004M04	188	G	0.88	A	0.13	0.22	CTCCATTTTCCCTAARGCTGCCACTCTTGGG
A004M43	78	C	0.81	A	0.19	0.30	GCAGTTTTACTGTACMAGAAGTGCAATGCTA
A004N13	110	A	0.63	G	0.38	0.47	TATCCTTTCCTCTGCRGGGACTAAACAAGAA
A004N44	65	G	0.88	A	0.13	0.22	ACTCATTTAGCAAAARTCCTGAACACAATAT
A004P08	105	G	0.94	A	0.06	0.12	GGTACTTCCCTAGAGRGTCCCGAGGCTCAGA
A004Q09	25	T	0.63	G	0.38	0.47	CAATGATAATACACCKTTGGATAAGGGGGAT
A004Q11	40	T	0.56	C	0.44	0.49	TCAGCACAGGAATTGYAATCTTCTCACTTCA
A004R33	68	C	0.94	T	0.06	0.12	CCCAACTACGATAAGYCATTGCCGGATGCTG
A004R38	74	T	0.94	C	0.06	0.12	TTTTTCTGATATACTYCTGAAAATTTTATAA
A005C35	158	C	0.94	T	0.06	0.12	GGGGCCTTGTGTTCCYGCCATCGGACAGCTG
A006N42	138	G	0.81	A	0.19	0.30	GTACTGGAAGTGGGARAGGCAAGGTCTGCTA
A006O23	37	G	0.94	A	0.06	0.12	GGGTGTGAGAAGCACRCAATAGGAAGTCTCT
A006P16	33	T	0.88	C	0.13	0.22	TTGTTCAGGCTGATCYAAACTCCTAGGCTCA
A006P20	149	A	0.44	G	0.56	0.49	ATCCTTTCCCTGCTARAAAGACAAAACAAAA
A006Q32	19	G	0.13	A	0.88	0.22	TTCATTGGCATTAAGRCATTCACAATGCTGT
A006Q32	84	G	0.81	A	0.19	0.30	TTTTCTTCATCGCTARAAGGAGTAATCCTTT
A006Q33	86	C	0.94	A	0.06	0.12	TGTCCTTTCTCAATTMACAAATGCTGTTAAA
A006R10	61	T	0.88	C	0.13	0.22	TGTTCTGCTCATAATYCCAATATGTACCAGA
A006R44	78	A	0.38	G	0.63	0.47	GCCAACGTGCTGATCRGTGCCTGTCCTGGAG
A006T39	130	G	0.88	C	0.13	0.22	TTTTTATCCTGAAATSTTTTTAGAAGCCCTG
A006U19	46	G	0.94	A	0.06	0.12	TACTGGATAACACTTRTTGGCCCATGACCTC
A006U44	237	C	0.75	G	0.25	0.38	AGGACTATTTCCATGSATGTGTTATTGGCAG
A006X15	172	A	0.81	G	0.19	0.30	GACTGCTGCCCCCAGRCAGGCAGGGGGTGTG
A006Y09	47	C	0.25	T	0.75	0.38	GGCTGAAACAGTGCCYAAGCTGGTCAGAGAT
A006Y32	176	G	0.19	A	0.81	0.30	ATTCTTTTCTTCACCRTAAAGGCTGTTCTTG
A006Y36	72	T	0.69	C	0.31	0.43	TCCTTAATCTCAAAGYATTTTTAGTAATACA

A007B18	156	A	0.88	T	0.13	0.22	TCCCACGGTGGAATAWTACACACAATTACAC
A007B24	62	G	0.38	T	0.63	0.47	GGAACAGAATGACAGKGGGATGCTGAGGAGC
A007C36	22	A	0.94	G	0.06	0.12	TTACTGATATTCATTRATTATTCCATAGGAC
A007C36	49	T	0.94	A	0.06	0.12	AGGACAGTTGTTTGAWTTGGTGCCACCTTAT
A007C36	67	G	0.94	T	0.06	0.12	TGGTGCCACCTTATTKCCCCTTTATACAGAT
A007D14	54	A	0.50	G	0.50	0.50	AAAGTTAAAAGGATARCGGTTACAGGAAAGT
A007D35	53	G	0.81	C	0.19	0.30	ATGTCTTGAGAACATSAAATGAATTGGACAA
A007E33	36	T	0.88	A	0.13	0.22	CACCTTCAAAAATTAWTGTGACTTACGGAAA
A007G47	40	A	0.94	G	0.06	0.12	TACCAGGCAAATAATRGTACATCCCCAAACC
A007H07	180	T	0.94	C	0.06	0.12	TGCCTACCATCTTCAYGGCCTCTGGGCACAA
A007I32	134	T	0.94	C	0.06	0.12	GGGGCGCTCGGGAGAYTGTGGACAATACCAA
A007K44	103	T	0.88	C	0.13	0.22	TTTATTATTATTATTYTGAGATAAGGTCATG
A007L07	150	T	0.69	G	0.31	0.43	CGCTGGGTGTGGGTTKATTCAGAGGCCCACA
A008B14	99	C	0.94	T	0.06	0.12	GATTCTACAGCACGCYGACACTAACACATCA
A008B43	93	A	0.88	G	0.13	0.22	TGTGCCAACTCAAGGRGCTACCTTGACATTA
A008C11	110	G	0.94	A	0.06	0.12	GCTCGTTCTGCAGGARTGGTGGTGGAAGGCC
A008C11	213	T	0.13	C	0.88	0.22	ATGGCGGTGGTGGCAYGGGAGCCTATGCCCC
A008C18	57	A	0.88	G	0.13	0.22	TCTAGAAATGTCTAARAACAACCTTTTTTAT
stSG8656	44	C	0.94	T	0.06	0.12	ATGACCTTGATGCCGYGGAATTATATTCAGA
stSG8880	28	C	0.94	T	0.06	0.12	CTGTACCCCCGACGTYTCCCCTGCTCGGCAC
stSG8904	35	G	0.88	A	0.13	0.22	TCACGCTGATCCAGCRGGCACCCTGCTTAAG
stSG8917	64	G	0.75	A	0.25	0.38	GTAACTATGACTAGARAGGCAGAGGAGTGGG
stSG8944	30	C	0.44	T	0.56	0.49	TTGTAAGGATGTTTCYATAGAAATCACGGAT
stSG8944	48	T	0.69	C	0.31	0.43	TAGAAATCACGGATAYATCACCAGTCTACAG
stSG8944	59	T	0.38	C	0.63	0.47	GATAGTATCACCAGTYACAGCCACTATCTAT
stSG90	40	A	0.25	G	0.75	0.38	CGAGGAGTAGCCAGGRGGCGAGACACAAAAA
stSG90	69	G	0.25	C	0.75	0.38	AAAGGCCTGGGACAGSTCAGTACAAGTCAGG
stSG9044	67	G	0.56	A	0.44	0.49	TGACTGTAGAGGATTRATGATCCCTGAATAC
stSG9062	83	C	0.38	G	0.63	0.47	GTACAGCAGGCTCTASCATTCTCTCTCTCTT
stSG9073	88	G	0.75	A	0.25	0.38	CTGGGCATGGCCGTGRCACCCTGTGTGGCGA
stSG9075	65	C	0.94	T	0.06	0.12	GATTCTACAGCACGCYGACACTAACACATCA
stSG9354	41	C	0.25	T	0.75	0.38	CCAGGGAGACACCTCYGTGAGATGACCTGCA
stSG9535	42	C	0.75	T	0.25	0.38	CTCCCTGCCAGTCTTYCCGTCTAACCCTCAG
stSG9615	38	A	0.56	T	0.44	0.49	GGTTGGAATGTGTATWCAACTTGATGATGAA
stSG9615	156	A	0.56	G	0.44	0.49	AATAAGTGTTGTGAARGATTTTATTATAAAT
stSG9673	82	A	0.88	G	0.13	0.22	TCCCTCTAATGAAGGRAAGGGTTTTTGAACA
stSG9757	195	G	0.94	T	0.06	0.12	TTAATCATTACTATTKCAACTCCGTATTTTC
stD22S972E	20	A	0.56	G	0.44	0.49	TCCAGGAGCTGTTATRCTACCAGTTTCTGGC
stSG10082	48	G	0.88	A	0.13	0.22	ACTGGCAGGGATTTGRTATCTAAACATAGAA
stSG10082	58	A	0.75	C	0.25	0.38	ATTTGCGTATCTAAAMTAGAAAAGGTACAGT
stSG1398	73	T	0.81	A	0.19	0.30	TTGCTTTTTATAATTWAAAGCAAATAACACA
stSG1437	71	G	0.25	T	0.75	0.38	AAGTTTTGACTTTGGKTCAAGTTTTTATTAC
stSG1446	106	T	0.50	A	0.50	0.50	TCAGCGTGAGATGATWTTGATTAAACTTGCT
stSG1446	147	G	0.75	C	0.25	0.38	TAGTCAAACGTCGAASTTGCTTGAGATGGCT
stSG149	107	G	0.19	T	0.81	0.30	TTGGGGAGGAACCATKCTCCNTCTGGGCCGC
stSG1514	78	T	0.81	G	0.19	0.30	TGGGTTTCTGTGGGAKCAGCGGGGGCCTCCT
stSG9800	134	C	0.50	A	0.50	0.50	TTAGTTTGGATTAATMGACTTAAGAAAACAA
stSG9828	32	G	0.88	A	0.13	0.22	ATTATGTGTTTCAGARTTATTAAAAAGGCTA
stSG9889	128	C	0.94	A	0.06	0.12	AGGAACTGAGAAAGAMCTGCCTAAGCAGCAC
stSG9950	139	G	0.88	A	0.13	0.22	AAAATACTTGGTTAARTTGAAAGGACCTAGT
stSG9961	33	A	0.19	G	0.81	0.30	TCTATTAGATAAAATRCAGATAAAGAATCTG
stSG9961	45	T	0.63	C	0.38	0.47	AATAACAGATAAAGAYCTGGAGAAAGGCTTT
stVPREB	30	G	0.94	A	0.06	0.12	ATATTTCTCACAATCRACAAGAGCCAGGGCC
stSG1615	57	T	0.58	C	0.42	0.49	GAGACATCCAGCCCAYTCTCTGGAACAGGAA
stSG1615	79	T	0.75	C	0.25	0.38	GGAACAGGAAAGATGYCGGGGAGGGAACACA
stSG1615	88	G	0.42	A	0.58	0.49	AAGATGATCGGGGAGRAACACAGGTCAGTNT
stSG1615	119	G	0.50	A	0.50	0.50	TGGGGACAGGGGTCARGTGGACACGGGGGTG
stSG1813	41	C	0.50	T	0.50	0.50	GTGAGGGCCCAGGGTYTCCACGGAGAGGACA
stSG1828	191	G	0.50	A	0.50	0.50	TGCTGTAGCCAAATTRTTGTCATACCAGGAA
stSG2020	51	C	0.75	T	0.25	0.38	ATTAGAAAAGGACGCYCTGTTGGCTGAACAA

stSG2125	55	A	0.83	G	0.17	0.28	TTTACAAAATTTCATRGAACTGACAATGTTA
stSG2294	139	T	0.92	C	0.08	0.15	AACACTGCAAAAACCYTCAAGCATAAAAAAG
stSG2314	89	T	0.75	A	0.25	0.38	ATGTCTTTTCCCAGTWGTCATATTTTTGTCC
stSG2417	84	T	0.83	C	0.17	0.28	ACTCTCTTATGACAAYAGTGATTGANCTCTA
stSG2482	121	A	0.08	T	0.92	0.15	TGANGCAGGCTATGGWTAAAAGAAACAACAA
stSG2623	77	C	0.92	T	0.08	0.15	TTGTCTTTTTTTTTCYGGCAAACTTTCTGCT
stSG2679	39	A	0.58	G	0.42	0.49	TACATTAATTTTCTTRTGAACACAGTAGACA
stSG2773	49	C	0.83	T	0.17	0.28	ATATACACTGTTTATYTTTTTCTTTTTCACG
stSG3009	88	C	0.92	T	0.08	0.15	TTACTTTTTATGTAGYTAAGTGTGTTTATAA
stSG3094	79	C	0.75	G	0.25	0.38	CTCCCCAGAGTAAAASGTTTTCTCTGGGNCT
stSG3234	74	C	0.94	G	0.06	0.12	TCTNCAGATTGCACTSTAAGATCCTAGTTAC
stSG3248	40	A	0.38	G	0.63	0.47	ACATTCAGAATTATGRAAACAATTAGTTATA
stSG3277	43	A	0.75	G	0.25	0.38	TCATTTGCTACATGARCAGAGGCAGAGTATT
stSG3349	141	T	0.31	A	0.69	0.43	CCTTTTAAAAAATGTWGAAATTTTAAGTGGG
stSG3388	28	T	0.94	C	0.06	0.12	AGTGAATTAGGGAGTYCTTGTTGACCCCTTT
stSG3552	40	G	0.56	A	0.44	0.49	AAAACCACATGTNCTRTAAGTGGGAGATAAA
stSG3809	87	T	0.44	C	0.56	0.49	AGTTACAGCCCCCTCYCACTCCTGTATCTGC
stSG3809	122	G	0.63	T	0.38	0.47	GGGTGGTGATGTGTTKGCTCCTAGACTCTCT
stSG3809	123	G	0.88	C	0.13	0.22	GGTGGTGATGTGTTTSCTCCTAGACTCTCTC
stSG3885	36	G	0.06	C	0.94	0.12	ATTTCTGACATTCATSCCAAAGANGGCAAAG
stSG3927	84	A	0.94	C	0.06	0.12	ACAAAATAAACCGCTMGTTTTTCTGGCTCCA
stSG3927	118	T	0.00	C	1.00	0.00	CACGCCATATGAAGCYGCCAATGTCACTTAT
stSG4025	41	G	0.88	A	0.13	0.22	ATCAACAGCTGCTACRTTCACCCCAGAGGTG
stSG4044	22	A	0.44	G	0.56	0.49	TAATATGGGGGGTCTRAACACAGCACCCCCA
stSG4085	30	A	0.94	C	0.06	0.12	GCCCCAGTGATTCTCMTACATTTTTCACCTC
stSG4085	97	C	0.69	T	0.31	0.43	TTTTCTTGCCTGGAGYTTCATTGTTCACCCT
stSG4148	68	T	0.38	A	0.63	0.47	GATAAGCAGATCAGCWGCCAGCCAAGCTCAT
stSG4389	52	G	0.38	T	0.63	0.47	GGCAGTATTTTAAAAKATTTCTCTAATGTTT
stSG4494	71	T	0.94	C	0.06	0.12	ATTATTTCAGTCATCYTAACATGTGACTTTA
stSG4537	42	G	0.94	T	0.06	0.12	CCTCTGGCGAGCCCTKCGGCTCCACATCCTC
stSG4702	124	C	0.94	T	0.06	0.12	ACAGTTGCTGACTCCYGGCCTGTGGGCAGNC
stSG4978	102	C	0.44	G	0.56	0.49	AGAGAGGCATCACTGSGCTGCATCTGCCATG
stSG6328	117	G	0.50	C	0.50	0.50	GCTTTAACAGAAACTSAACTCTTCACGCTTG
stSG8971	95	T	0.88	C	0.13	0.22	AGCAATTTAATACAGYGAAAACAAATACAAT
A002Q12	26	T	0.25	C	0.75	0.38	TCATTATTCTCTCCTYAGATTATTAAATATT
A002Q19	32	C	0.75	G	0.25	0.38	TCCTTCCCCTCCTGCSCCACACTGCTGGCCA
A002Q20	138	T	0.88	C	0.13	0.22	GCCTGCATTTGGCTTYGTGCCTGAAAAAGAA
A002S01	83	A	0.69	C	0.31	0.43	AAATGAAGATTTAATMTTCCTAAATTTAAGT
A002T26	86	C	0.81	T	0.19	0.30	CGTCAAAAGAAAACCYCCCCGGGACCCACTGT
A002V42	50	T	0.06	C	0.94	0.12	TTTCATTCAGTTATAYCTTTGGCTCAGCTAG
A002Y34	89	A	0.88	G	0.13	0.22	TAACAGAAACGCCTTRGACACTATGTTTGGG
A002Y45	85	C	0.75	A	0.25	0.38	GTGTGTCAGGATGCAMTGAAAGCCCTCGGCT
A002Y45	106	G	0.38	C	0.63	0.47	AGCCCTCGGCTCGGTSTTAGCCAATCTTCCT
A003B21	49	T	0.63	C	0.38	0.47	GACAACTTAAAACTTYTAGTGACATTGCTGT
A003B21	120	T	0.63	A	0.38	0.47	TTAAAAGAGCAAAGTWCCCCTCCCTTTCTTA
A003B29	68	G	0.88	A	0.13	0.22	TTTTGGCCATAGACARTTATTTTGATTCTAA
stSG9569	191	A	0.19	T	0.81	0.30	ATATGTATATATATAWTTTTTTTAATTCCTC
stSG9574	43	T	0.81	G	0.19	0.30	TTGGGGCAAAAGAGTKTCTTCATTATCAATC
stSG9792	105	G	0.75	T	0.25	0.38	CTGGTGCGCTGAGGCKTACACACCGGCAGAA
stSG9792	108	C	0.94	T	0.06	0.12	GTGCGCTGAGGCTGTYACACCGGCAGAACAG
stSG9915	81	T	0.94	C	0.06	0.12	CAAACCCATTTAAGTYGGAATGATTTATATG
stSG9997	99	C	0.88	G	0.13	0.22	GCCCTAATAATCCAGSATTCCTNACTCTCTT
A004A22	125	G	0.94	A	0.06	0.12	TATCTGGCGAGGAGGRCGGCATGGAGTCCAG
A004A30	135	G	0.94	C	0.06	0.12	GAATTTTAGATGCAGSATCATTTTATATATA
A004B17	146	T	0.25	C	0.75	0.38	ATTGCTGGGGCTCTAYTCCACAATTTGTTTT
A004B36	107	A	0.94	G	0.06	0.12	TTGGAGTGCACTGGCRTCCCTCAGATTTGTC
A004B39	58	G	0.94	T	0.06	0.12	CCTCCCCTCCAGACCKCTCCTTCTCCCTGCT
A004F06	71	C	0.94	T	0.06	0.12	ATAATTTATACCACAYCTGAAGAAATTATCT
A004F17	47	G	0.94	A	0.06	0.12	TATGGACTATGTACARACAATACAAGAGGCG
A004G25	85	C	0.94	T	0.06	0.12	AATCATTCTCTCTTCYTTCACATGGTGTACT
A004H43	35	C	0.75	T	0.25	0.38	GGACTTCCTAGCCTTYAGCAAGCTTAGAGGA

A004H45	26	T	0.75	G	0.25	0.38	GAAGTTGCTATAGGTKCTCTTTCTCAAAAGT
A004I05	49	G	0.06	A	0.94	0.12	TAATGCCATTTTGATRTTAACATTACGTGTC
A004I26	62	A	0.94	G	0.06	0.12	AAAAAGATTAAAAACRAAATAATATTAAAGG
A004I35	45	C	0.13	T	0.88	0.22	CCCCAGGTTAAACACYTGTAATTTACCTTGA
A004I36	173	A	0.63	C	0.38	0.47	AGACAATGTCACTTGMAACACAAGGTATGAA
A004I36	190	T	0.31	C	0.69	0.43	AACACAAGGTATGAAYATAAATAATAGTCAG
A004M04	188	G	0.88	A	0.13	0.22	CTCCATTTTCCCTAARGCTGCCACTCTTGGG
A004M43	78	C	0.81	A	0.19	0.30	GCAGTTTTACTGTACMAGAAGTGCAATGCTA
A004N13	110	A	0.63	G	0.38	0.47	TATCCTTTCCTCTGCRGGGACTAAACAAGAA
A004N44	65	G	0.88	A	0.13	0.22	ACTCATTTAGCAAAARTCCTGAACACAATAT
A004P08	105	G	0.94	A	0.06	0.12	GGTACTTCCCTAGAGRGTCCCGAGGCTCAGA
A004Q09	25	T	0.63	G	0.38	0.47	CAATGATAATACACCKTTGGATAAGGGGGAT
A004Q11	40	T	0.56	C	0.44	0.49	TCAGCACAGGAATTGYAATCTTCTCACTTCA
A004R33	68	C	0.94	T	0.06	0.12	CCCAACTACGATAAGYCATTGCCGGATGCTG
A004R38	74	T	0.94	C	0.06	0.12	TTTTTCTGATATACTYCTGAAAATTTTATAA
A005C35	158	C	0.94	T	0.06	0.12	GGGGCCTTGTGTTCCYGCCATCGGACAGCTG
A006N42	138	G	0.81	A	0.19	0.30	GTACTGGAAGTGGGARAGGCAAGGTCTGCTA
A006O23	37	G	0.94	A	0.06	0.12	GGGTGTGAGAAGCACRCAATAGGAAGTCTCT
A006P16	33	T	0.88	C	0.13	0.22	TTGTTCAGGCTGATCYAAACTCCTAGGCTCA
A006P20	149	A	0.44	G	0.56	0.49	ATCCTTTCCCTGCTARAAAGACAAAACAAAA
A006Q32	19	G	0.13	A	0.88	0.22	TTCATTGGCATTAAGRCATTCACAATGCTGT
A006Q32	84	G	0.81	A	0.19	0.30	TTTTCTTCATCGCTARAAGGAGTAATCCTTT
A006Q33	86	C	0.94	A	0.06	0.12	TGTCCTTTCTCAATTMACAAATGCTGTTAAA
A006R10	61	T	0.88	C	0.13	0.22	TGTTCTGCTCATAATYCCAATATGTACCAGA
A006R44	78	A	0.38	G	0.63	0.47	GCCAACGTGCTGATCRGTGCCTGTCCTGGAG
A006T39	130	G	0.88	C	0.13	0.22	TTTTTATCCTGAAATSTTTTTAGAAGCCCTG
A006U19	46	G	0.94	A	0.06	0.12	TACTGGATAACACTTRTTGGCCCATGACCTC
A006U44	237	C	0.75	G	0.25	0.38	AGGACTATTTCCATGSATGTGTTATTGGCAG
A006X15	172	A	0.81	G	0.19	0.30	GACTGCTGCCCCCAGRCAGGCAGGGGGTGTG
A006Y09	47	C	0.25	T	0.75	0.38	GGCTGAAACAGTGCCYAAGCTGGTCAGAGAT
A006Y32	176	G	0.19	A	0.81	0.30	ATTCTTTTCTTCACCRTAAAGGCTGTTCTTG
A006Y36	72	T	0.69	C	0.31	0.43	TCCTTAATCTCAAAGYATTTTTAGTAATACA
A007B18	156	A	0.88	T	0.13	0.22	TCCCACGGTGGAATAWTACACACAATTACAC
A007B24	62	G	0.38	T	0.63	0.47	GGAACAGAATGACAGKGGGATGCTGAGGAGC
A007C36	22	A	0.94	G	0.06	0.12	TTACTGATATTCATTRATTATTCCATAGGAC
A007C36	49	T	0.94	A	0.06	0.12	AGGACAGTTGTTTGAWTTGGTGCCACCTTAT
A007C36	67	G	0.94	T	0.06	0.12	TGGTGCCACCTTATTKCCCCTTTATACAGAT
A007D14	54	A	0.50	G	0.50	0.50	AAAGTTAAAAGGATARCGGTTACAGGAAAGT
A007D35	53	G	0.81	C	0.19	0.30	ATGTCTTGAGAACATSAAATGAATTGGACAA
A007E33	36	T	0.88	A	0.13	0.22	CACCTTCAAAAATTAWTGTGACTTACGGAAA
A007G47	40	A	0.94	G	0.06	0.12	TACCAGGCAAATAATRGTACATCCCCAAACC
A007H07	180	T	0.94	C	0.06	0.12	TGCCTACCATCTTCAYGGCCTCTGGGCACAA
A007I32	134	T	0.94	C	0.06	0.12	GGGGCGCTCGGGAGAYTGTGGACAATACCAA
A007K44	103	T	0.88	C	0.13	0.22	TTTATTATTATTATTYTGAGATAAGGTCATG
A007L07	150	T	0.69	G	0.31	0.43	CGCTGGGTGTGGGTTKATTCAGAGGCCCACA
A008B14	99	C	0.94	T	0.06	0.12	GATTCTACAGCACGCYGACACTAACACATCA
A008B43	93	A	0.88	G	0.13	0.22	TGTGCCAACTCAAGGRGCTACCTTGACATTA
A008C11	110	G	0.94	A	0.06	0.12	GCTCGTTCTGCAGGARTGGTGGTGGAAGGCC
A008C11	213	T	0.13	C	0.88	0.22	ATGGCGGTGGTGGCAYGGGAGCCTATGCCCC
A008C18	57	A	0.88	G	0.13	0.22	TCTAGAAATGTCTAARAACAACCTTTTTTAT

Fragments prefaced stSG are from the Sanger Centre, UK.
Fragments prefaced with A are from Genethon, France.
Fragments without a prefix are from the Whitehead Institute.

Analysis of Polymorphisms

A. Preparation of Samples

Polymorphisms are detected in a target nucleic acid from an individual being analyzed. For assay of genomic DNA, virtually any biological sample (other than pure red blood cells) is suitable. For example, convenient tissue samples include whole blood, semen, saliva, tears, urine, fecal material, sweat, buccal, skin and hair. For assay of cDNA or mRNA, the tissue sample must be obtained from an organ in which the target nucleic acid is expressed. For example, if the target nucleic acid is a cytochrome P450, the liver is a suitable source.

Many of the methods described below require amplification of DNA from target samples. This can be accomplished by e.g., PCR. *See generally PCR Technology: Principles and Applications for DNA Amplification* (ed. H.A. Erlich, Freeman Press, NY, NY, 1992); *PCR Protocols: A Guide to Methods and Applications* (eds. Innis, et al., Academic Press, San Diego, CA, 1990); Mattila et al., *Nucleic Acids Res.* 19, 4967 (1991); Eckert et al., *PCR Methods and Applications* 1, 17 (1991); *PCR* (eds. McPherson et al., IRL Press, Oxford); and U.S. Patent 4,683,202 (each of which is incorporated by reference for all purposes).

Other suitable amplification methods include the ligase chain reaction (LCR) (see Wu and Wallace, *Genomics* 4, 560 (1989), Landegren et al., *Science* 241, 1077 (1988), transcription amplification (Kwoh et al., *Proc. Natl. Acad. Sci. USA* 86, 1173 (1989)), and self-sustained sequence replication (Guatelli et al., *Proc. Nat. Acad. Sci. USA,* 87, 1874 (1990)) and nucleic acid based sequence amplification (NASBA). The latter two amplification methods involve isothermal reactions based

on isothermal transcription, which produce both single stranded RNA (ssRNA) and double stranded DNA (dsDNA) as the amplification products in a ratio of about 30 or 100 to 1, respectively.

B. Detection of Polymorphisms in Target DNA

There are two distinct types of analysis depending whether a polymorphism in question has already been characterized. The first type of analysis is sometimes referred to as de novo characterization. This analysis compares target sequences in different individuals to identify points of variation, i.e., polymorphic sites. By analyzing a groups of individuals representing the greatest ethnic diversity among humans and greatest breed and species variety in plants and animals, patterns characteristic of the most common alleles/haplotypes of the locus can be identified, and the frequencies of such populations in the population determined. Additional allelic frequencies can be determined for subpopulations characterized by criteria such as geography, race, or gender. The de novo identification of the polymorphisms of the invention is described in the Examples section. The second type of analysis is determining which form(s) of a characterized polymorphism are present in individuals under test. There are a variety of suitable procedures, which are discussed in turn.

1. Allele-Specific Probes

The design and use of allele-specific probes for analyzing polymorphisms is described by e.g., Saiki et al., *Nature* 324, 163-166 (1986); Dattagupta, EP 235,726, Saiki, WO 89/11548. Allele-specific probes can be designed that hybridize to a segment of target DNA from one individual but do not hybridize to the corresponding segment from another individual due to the presence of different polymorphic forms in the respective segments from the two individuals. Hybridization conditions

should be sufficiently stringent that there is a significant difference in hybridization intensity between alleles, and preferably an essentially binary response, whereby a probe hybridizes to only one of the alleles. Some probes are designed to hybridize to a segment of target DNA such that the polymorphic site aligns with a central position (e.g., in a 15 mer at the 7 position; in a 16 mer, at either the 8 or 9 position) of the probe. This design of probe achieves good discrimination in hybridization between different allelic forms.

Allele-specific probes are often used in pairs, one member of a pair showing a perfect match to a reference form of a target sequence and the other member showing a perfect match to a variant form. Several pairs of probes can then be immobilized on the same support for simultaneous analysis of multiple polymorphisms within the same target sequence.

2. Tiling Arrays

The polymorphisms can also be identified by hybridization to nucleic acid arrays, some example of which are described by WO 95/11995 (incorporated by reference in its entirety for all purposes). One form of such arrays is described in the Examples section in connection with de novo identification of polymorphisms. The same array or a different array can be used for analysis of characterized polymorphisms. WO 95/11995 also describes subarrays that are optimized for detection of a variant forms of a precharacterized polymorphism. Such a subarray contains probes designed to be complementary to a second reference sequence, which is an allelic variant of the first reference sequence. The second group of probes is designed by the same principles as described in the Examples except that the probes exhibit complementarily to the second reference sequence. The inclusion of a second group (or further groups) can be particular useful for analyzing short subsequences of

the primary reference sequence in which multiple mutations are expected to occur within a short distance commensurate with the length of the probes (*i.e.*, two or more mutations within 9 to 21 bases).

3. Allele-Specific Primers

An allele-specific primer hybridizes to a site on target DNA overlapping a polymorphism and only primes amplification of an allelic form to which the primer exhibits perfect complementarily. See Gibbs, *Nucleic Acid Res.* 17, 2427-2448 (1989). This primer is used in conjunction with a second primer which hybridizes at a distal site. Amplification proceeds from the two primers leading to a detectable product signifying the particular allelic form is present. A control is usually performed with a second pair of primers, one of which shows a single base mismatch at the polymorphic site and the other of which exhibits perfect complementarily to a distal site. The single-base mismatch prevents amplification and no detectable product is formed. The method works best when the mismatch is included in the 3'-most position of the oligonucleotide aligned with the polymorphism because this position is most destabilizing to elongation from the primer. *See, e.g.*, WO 93/22456.

4. Direct-Sequencing

The direct analysis of the sequence of polymorphisms of the present invention can be accomplished using either the dideoxy chain termination method or the Maxam Gilbert method (see Sambrook et al., *Molecular Cloning, A Laboratory Manual* (2nd Ed., CSHP, New York 1989); Zyskind et al., *Recombinant DNA Laboratory Manual*, (Acad. Press, 1988)).

5. Denaturing Gradient Gel Electrophoresis

Amplification products generated using the polymerase chain reaction can be analyzed by the use of denaturing gradient gel electrophoresis. Different alleles can be identified based on the different sequence-dependent melting properties and electrophoretic migration of DNA in solution. Erlich, ed., *PCR Technology, Principles and Applications for DNA Amplification*, (W.H. Freeman and Co, New York, 1992), Chapter 7.

6. Single-Strand Conformation Polymorphism Analysis

Alleles of target sequences can be differentiated using single-strand conformation polymorphism analysis, which identifies base differences by alteration in electrophoretic migration of single stranded PCR products, as described in Orita et al., *Proc. Nat. Acad. Sci.* 86, 2766-2770 (1989). Amplified PCR products can be generated as described above, and heated or otherwise denatured, to form single stranded amplification products. Single-stranded nucleic acids may refold or form secondary structures which are partially dependent on the base sequence. The different electrophoretic mobilities of single-stranded amplification products can be related to base-sequence difference between alleles of target sequences.

III. Methods of Use

After determining polymorphic form(s) present in an individual at one or more polymorphic sites, this information can be used in a number of methods.

A. Forensics

Determination of which polymorphic forms occupy a set of polymorphic sites in an individual identifies a

set of polymorphic forms that distinguishes the individual. *See generally* National Research Council, *The Evaluation of Forensic DNA Evidence* (Eds. Pollard et al., National Academy Press, DC, 1996). The more sites that are analyzed the lower the probability that the set of polymorphic forms in one individual is the same as that in an unrelated individual. Preferably, if multiple sites are analyzed, the sites are unlinked. Thus, polymorphisms of the invention are often used in conjunction with polymorphisms in distal genes. Preferred polymorphisms for use in forensics are diallelic because the population frequencies of two polymorphic forms can usually be determined with greater accuracy than those of multiple polymorphic forms at multi-allelic loci.

The capacity to identify a distinguishing or unique set of forensic markers in an individual is useful for forensic analysis. For example, one can determine whether a blood sample from a suspect matches a blood or other tissue sample from a crime scene by determining whether the set of polymorphic forms occupying selected polymorphic sites is the same in the suspect and the sample. If the set of polymorphic markers does not match between a suspect and a sample, it can be concluded (barring experimental error) that the suspect was not the source of the sample. If the set of markers does match, one can conclude that the DNA from the suspect is consistent with that found at the crime scene. If frequencies of the polymorphic forms at the loci tested have been determined (e.g., by analysis of a suitable population of individuals), one can perform a statistical analysis to determine the probability that a match of suspect and crime scene sample would occur by chance.

p(ID) is the probability that two random individuals have the same polymorphic or allelic form at a given polymorphic site. In diallelic loci, four

genotypes are possible: AA, AB, BA, and BB. If alleles A and B occur in a haploid genome of the organism with frequencies x and y, the probability of each genotype in a diploid organism are (see WO 95/12607):

Homozygote: $p(AA) = x^2$
Homozygote: $p(BB) = y^2 = (1-x)^2$
Single Heterozygote: $p(AB) = p(BA) = xy = x(1-x)$
Both Heterozygotes: $p(AB+BA) = 2xy = 2x(1-x)$

The probability of identity at one locus (i.e, the probability that two individuals, picked at random from a population will have identical polymorphic forms at a given locus) is given by the equation:

$p(ID) = (x^2)^2 + (2xy)^2 + (y^2)^2$.

These calculations can be extended for any number of polymorphic forms at a given locus. For example, the probability of identity p(ID) for a 3-allele system where the alleles have the frequencies in the population of x, y and z, respectively, is equal to the sum of the squares of the genotype frequencies:

$p(ID) = x^4 + (2xy)^2 + (2yz)^2 + (2xz)^2 + z^4 + y^4$

In a locus of n alleles, the appropriate binomial expansion is used to calculate p(ID) and p(exc).

The cumulative probability of identity (cum p(ID)) for each of multiple unlinked loci is determined by multiplying the probabilities provided by each locus.

cum p(ID) = p(ID1)p(ID2)p(ID3).... p(IDn)

The cumulative probability of non-identity for n loci (i.e. the probability that two random individuals will be different at 1 or more loci) is given by the equation:

cum p(nonID) = 1-cum p(ID).

If several polymorphic loci are tested, the cumulative probability of non-identity for random individuals becomes very high (e.g., one billion to one). Such probabilities can be taken into account together with other evidence in determining the guilt or innocence of the suspect.

B. Paternity Testing

The object of paternity testing is usually to determine whether a male is the father of a child. In most cases, the mother of the child is known and thus, the mother's contribution to the child's genotype can be traced. Paternity testing investigates whether the part of the child's genotype not attributable to the mother is consistent with that of the putative father. Paternity testing can be performed by analyzing sets of polymorphisms in the putative father and the child.

If the set of polymorphisms in the child attributable to the father does not match the putative father, it can be concluded, barring experimental error, that the putative father is not the real father. If the set of polymorphisms in the child attributable to the father does match the set of polymorphisms of the putative father, a statistical calculation can be performed to determine the probability of coincidental match.

The probability of parentage exclusion (representing the probability that a random male will have a polymorphic form at a given polymorphic site that makes him incompatible as the father) is given by the equation (see WO 95/12607):

$$p(exc) = xy(1-xy)$$

where x and y are the population frequencies of alleles A and B of a diallelic polymorphic site.

(At a triallelic site $p(exc) = xy(1-xy) + yz(1-yz) + xz(1-xz) + 3xyz(1-xyz)$), where x, y and z and the respective population frequencies of alleles A, B and C).

The probability of non-exclusion is

$$p(non\text{-}exc) = 1-p(exc)$$

The cumulative probability of non-exclusion (representing the value obtained when n loci are used) is thus:

$$cum\ p(non\text{-}exc) = p(non\text{-}exc1)p(non\text{-}exc2)p(non\text{-}exc3)....p(non\text{-}excn)$$

The cumulative probability of exclusion for n loci (representing the probability that a random male will be excluded)

cum p(exc) = 1 - cum p(non-exc).

If several polymorphic loci are included in the analysis, the cumulative probability of exclusion of a random male is very high. This probability can be taken into account in assessing the liability of a putative father whose polymorphic marker set matches the child's polymorphic marker set attributable to his/her father.

C. Correlation of Polymorphisms with Phenotypic Traits

The polymorphisms of the invention may contribute to the phenotype of an organism in different ways. Some polymorphisms occur within a protein coding sequence and contribute to phenotype by affecting protein structure. The effect may be neutral, beneficial or detrimental, or both beneficial and detrimental, depending on the circumstances. For example, a heterozygous sickle cell mutation confers resistance to malaria, but a homozygous sickle cell mutation is usually lethal. Other polymorphisms occur in noncoding regions but may exert phenotypic effects indirectly via influence on replication, transcription, and translation. A single polymorphism may affect more than one phenotypic trait. Likewise, a single phenotypic trait may be affected by polymorphisms in different genes. Further, some polymorphisms predispose an individual to a distinct mutation that is causally related to a certain phenotype.

Phenotypic traits include diseases that have known but hitherto unmapped genetic components (e.g., agammaglobulimenia, diabetes insipidus, Lesch-Nyhan syndrome, muscular dystrophy, Wiskott-Aldrich syndrome, Fabry's disease, familial hypercholesterolemia, polycystic kidney disease, hereditary spherocytosis, von

Willebrand's disease, tuberous sclerosis, hereditary hemorrhagic telangiectasia, familial colonic polyposis, Ehlers-Danlos syndrome, osteogenesis imperfecta, and acute intermittent porphyria). Phenotypic traits also include symptoms of, or susceptibility to, multifactorial diseases of which a component is or may be genetic, such as autoimmune diseases, inflammation, cancer, diseases of the nervous system, and infection by pathogenic microorganisms. Some examples of autoimmune diseases include rheumatoid arthritis, multiple sclerosis, diabetes (insulin-dependent and non-independent), systemic lupus erythematosus and Graves disease. Some examples of cancers include cancers of the bladder, brain, breast, colon, esophagus, kidney, leukemia, liver, lung, oral cavity, ovary, pancreas, prostate, skin, stomach and uterus. Phenotypic traits also include characteristics such as longevity, appearance (e.g., baldness, obesity), strength, speed, endurance, fertility, and susceptibility or receptivity to particular drugs or therapeutic treatments.

Correlation is performed for a population of individuals who have been tested for the presence or absence of a phenotypic trait of interest and for polymorphic markers sets. To perform such analysis, the presence or absence of a set of polymorphisms (i.e. a polymorphic set) is determined for a set of the individuals, some of whom exhibit a particular trait, and some of which exhibit lack of the trait. The alleles of each polymorphism of the set are then reviewed to determine whether the presence or absence of a particular allele is associated with the trait of interest. Correlation can be performed by standard statistical methods such as a κ-squared test and statistically significant correlations between polymorphic form(s) and phenotypic characteristics are noted. For example, it might be found that the presence of allele A1 at polymorphism A correlates with heart disease. As a

further example, it might be found that the combined presence of allele A1 at polymorphism A and allele B1 at polymorphism B correlates with increased milk production of a farm animal.

Such correlations can be exploited in several ways. In the case of a strong correlation between a set of one or more polymorphic forms and a disease for which treatment is available, detection of the polymorphic form set in a human or animal patient may justify immediate administration of treatment, or at least the institution of regular monitoring of the patient. Detection of a polymorphic form correlated with serious disease in a couple contemplating a family may also be valuable to the couple in their reproductive decisions. For example, the female partner might elect to undergo in vitro fertilization to avoid the possibility of transmitting such a polymorphism from her husband to her offspring. In the case of a weaker, but still statistically significant correlation between a polymorphic set and human disease, immediate therapeutic intervention or monitoring may not be justified. Nevertheless, the patient can be motivated to begin simple life-style changes (e.g., diet, exercise) that can be accomplished at little cost to the patient but confer potential benefits in reducing the risk of conditions to which the patient may have increased susceptibility by virtue of variant alleles. Identification of a polymorphic set in a patient correlated with enhanced receptiveness to one of several treatment regimes for a disease indicates that this treatment regime should be followed.

For animals and plants, correlations between characteristics and phenotype are useful for breeding for desired characteristics. For example, Beitz et al., US 5,292,639 discuss use of bovine mitochondrial polymorphisms in a breeding program to improve milk production in cows. To evaluate the effect of mtDNA D-loop sequence polymorphism on milk production, each cow

was assigned a value of 1 if variant or 0 if wildtype with respect to a prototypical mitochondrial DNA sequence at each of 17 locations considered. Each production trait was analyzed individually with the following animal model:

$Y_{ijkpn} = \mu + YS_i + P_j + X_k + \beta_1 + \ldots \beta_{17} + PE_n + a_n + e_p$

where Y_{ijknp} is the milk, fat, fat percentage, SNF, SNF percentage, energy concentration, or lactation energy record; μ is an overall mean; YS_i is the effect common to all cows calving in year-season; X_k is the effect common to cows in either the high or average selection line; β_1 to β_{17} are the binomial regressions of production record on mtDNA D-loop sequence polymorphisms; PE_n is permanent environmental effect common to all records of cow n; a_n is effect of animal n and is composed of the additive genetic contribution of sire and dam breeding values and a Mendelian sampling effect; and e_p is a random residual. It was found that eleven of seventeen polymorphisms tested influenced at least one production trait. Bovines having the best polymorphic forms for milk production at these eleven loci are used as parents for breeding the next generation of the herd.

D. Genetic Mapping of Phenotypic Traits

The previous section concerns identifying correlations between phenotypic traits and polymorphisms that directly or indirectly contribute to those traits. The present section describes identification of a physical linkage between a genetic locus associated with a trait of interest and polymorphic markers that are not associated with the trait, but are in physical proximity with the genetic locus responsible for the trait and co-segregate with it. Such analysis is useful for mapping a genetic locus associated with a phenotypic trait to a chromosomal position, and thereby cloning gene(s) responsible for the trait. *See* Lander et al., *Proc. Natl. Acad. Sci. (USA)* 83, 7353-7357 (1986); Lander et

al., *Proc. Natl. Acad. Sci. (USA)* 84, 2363-2367 (1987); Donis-Keller et al., *Cell* 51, 319-337 (1987); Lander et al., *Genetics* 121, 185-199 (1989)). Genes localized by linkage can be cloned by a process known as directional cloning. *See* Wainwright, *Med. J. Australia* 159, 170-174 (1993); Collins, *Nature Genetics* 1, 3-6 (1992) (each of which is incorporated by reference in its entirety for all purposes).

Linkage studies are typically performed on members of a family. Available members of the family are characterized for the presence or absence of a phenotypic trait and for a set of polymorphic markers. The distribution of polymorphic markers in an informative meiosis is then analyzed to determine which polymorphic markers co-segregate with a phenotypic trait. *See, e.g.*, Kerem et al., *Science* 245, 1073-1080 (1989); Monaco et al., *Nature* 316, 842 (1985); Yamoka et al., *Neurology* 40, 222-226 (1990); Rossiter et al., *FASEB Journal* 5, 21-27 (1991).

Linkage is analyzed by calculation of LOD (log of the odds) values. A lod value is the relative likelihood of obtaining observed segregation data for a marker and a genetic locus when the two are located at a recombination fraction θ, versus the situation in which the two are not linked, and thus segregating independently (Thompson & Thompson, *Genetics in Medicine* (5th ed, W.B. Saunders Company, Philadelphia, 1991); Strachan, "Mapping the human genome" in *The Human Genome* (BIOS Scientific Publishers Ltd, Oxford), Chapter 4). A series of likelihood ratios are calculated at various recombination fractions (θ), ranging from $\theta = 0.0$ (coincident loci) to $\theta = 0.50$ (unlinked). Thus, the likelihood at a given value of θ is: probability of data if loci linked at θ to probability of data if loci unlinked. The computed likelihoods are usually expressed as the $\log_{10}$ of this

ratio (i.e., a lod score). For example, a lod score of 3 indicates 1000:1 odds against an apparent observed linkage being a coincidence. The use of logarithms allows data collected from different families to be combined by simple addition. Computer programs are available for the calculation of lod scores for differing values of θ (e.g., LIPED, MLINK (Lathrop, *Proc. Nat. Acad. Sci. (USA)* 81, 3443-3446 (1984)). For any particular lod score, a recombination fraction may be determined from mathematical tables. *See* Smith et al., *Mathematical tables for research workers in human genetics* (Churchill, London, 1961); Smith, *Ann. Hum. Genet.* 32, 127-150 (1968). The value of θ at which the lod score is the highest is considered to be the best estimate of the recombination fraction.

Positive lod score values suggest that the two loci are linked, whereas negative values suggest that linkage is less likely (at that value of θ) than the possibility that the two loci are unlinked. By convention, a combined lod score of +3 or greater (equivalent to greater than 1000:1 odds in favor of linkage) is considered definitive evidence that two loci are linked. Similarly, by convention, a negative lod score of -2 or less is taken as definitive evidence against linkage of the two loci being compared. Negative linkage data are useful in excluding a chromosome or a segment thereof from consideration. The search focuses on the remaining non-excluded chromosomal locations.

IV. Modified Polypeptides and Gene Sequences

The invention further provides variant forms of nucleic acids and corresponding proteins. The nucleic acids comprise one of the sequences described in Table 1, column 8, in which the polymorphic position is occupied by one of the alternative bases for that position. Some nucleic acid encode full-length variant forms of

proteins. Similarly, variant proteins have the prototypical amino acid sequences of encoded by nucleic acid sequence shown in Table 1, column 8, (read so as to be in-frame with the full-length coding sequence of which it is a component) except at an amino acid encoded by a codon including one of the polymorphic positions shown in the Table. That position is occupied by the amino acid coded by the corresponding codon in any of the alternative forms shown in the Table.

Variant genes can be expressed in an expression vector in which a variant gene is operably linked to a native or other promoter. Usually, the promoter is a eukaryotic promoter for expression in a mammalian cell. The transcription regulation sequences typically include a heterologous promoter and optionally an enhancer which is recognized by the host. The selection of an appropriate promoter, for example trp, lac, phage promoters, glycolytic enzyme promoters and tRNA promoters, depends on the host selected. Commercially available expression vectors can be used. Vectors can include host-recognized replication systems, amplifiable genes, selectable markers, host sequences useful for insertion into the host genome, and the like.

The means of introducing the expression construct into a host cell varies depending upon the particular construction and the target host. Suitable means include fusion, conjugation, transfection, transduction, electroporation or injection, as described in Sambrook, *supra*. A wide variety of host cells can be employed for expression of the variant gene, both prokaryotic and eukaryotic. Suitable host cells include bacteria such as *E. coli*, yeast, filamentous fungi, insect cells, mammalian cells, typically immortalized, *e.g.*, mouse, CHO, human and monkey cell lines and derivatives thereof. Preferred host cells are able to process the variant gene product to produce an appropriate mature polypeptide.

Processing includes glycosylation, ubiquitination, disulfide bond formation, general post-translational modification, and the like.

The protein may be isolated by conventional means of protein biochemistry and purification to obtain a substantially pure product, *i.e.*, 80, 95 or 99% free of cell component contaminants, as described in Jacoby, *Methods in Enzymology* Volume 104, Academic Press, New York (1984); Scopes, *Protein Purification, Principles and Practice*, 2nd Edition, Springer-Verlag, New York (1987); and Deutscher (ed), *Guide to Protein Purification, Methods in Enzymology*, Vol. 182 (1990). If the protein is secreted, it can be isolated from the supernatant in which the host cell is grown. If not secreted, the protein can be isolated from a lysate of the host cells.

The invention further provides transgenic nonhuman animals capable of expressing an exogenous variant gene and/or having one or both alleles of an endogenous variant gene inactivated. Expression of an exogenous variant gene is usually achieved by operably linking the gene to a promoter and optionally an enhancer, and microinjecting the construct into a zygote. *See* Hogan et al., "Manipulating the Mouse Embryo, A Laboratory Manual," Cold Spring Harbor Laboratory. Inactivation of endogenous variant genes can be achieved by forming a transgene in which a cloned variant gene is inactivated by insertion of a positive selection marker. *See* Capecchi, *Science* 244, 1288-1292 (1989). The transgene is then introduced into an embryonic stem cell, where it undergoes homologous recombination with an endogenous variant gene. Mice and other rodents are preferred animals. Such animals provide useful drug screening systems.

In addition to substantially full-length polypeptides expressed by variant genes, the present

invention includes biologically active fragments of the polypeptides, or analogs thereof, including organic molecules which simulate the interactions of the peptides. Biologically active fragments include any portion of the full-length polypeptide which confers a biological function on the variant gene product, including ligand binding, and antibody binding. Ligand binding includes binding by nucleic acids, proteins or polypeptides, small biologically active molecules, or large cellular structures.

Polyclonal and/or monoclonal antibodies that specifically bind to variant gene products but not to corresponding prototypical gene products are also provided. Antibodies can be made by injecting mice or other animals with the variant gene product or synthetic peptide fragments thereof. Monoclonal antibodies are screened as are described, for example, in Harlow & Lane, *Antibodies, A Laboratory Manual*, Cold Spring Harbor Press, New York (1988); Goding, *Monoclonal antibodies, Principles and Practice* (2d ed.) Academic Press, New York (1986). Monoclonal antibodies are tested for specific immunoreactivity with a variant gene product and lack of immunoreactivity to the corresponding prototypical gene product. These antibodies are useful in diagnostic assays for detection of the variant form, or as an active ingredient in a pharmaceutical composition.

V. Kits

The invention further provides kits comprising at least one allele-specific oligonucleotide as described above. Often, the kits contain one or more pairs of allele-specific oligonucleotides hybridizing to different forms of a polymorphism. In some kits, the allele-specific oligonucleotides are provided immobilized to a substrate. For example, the same substrate can comprise allele-specific oligonucleotide probes for detecting at

least 10, 100 or all of the polymorphisms shown in Table 1. Optional additional components of the kit include, for example, restriction enzymes, reverse-transcriptase or polymerase, the substrate nucleoside triphosphates, means used to label (for example, an avidin-enzyme conjugate and enzyme substrate and chromogen if the label is biotin), and the appropriate buffers for reverse transcription, PCR, or hybridization reactions. Usually, the kit also contains instructions for carrying out the methods.

VI. Computer Systems For Storing Polymorphism Data

Fig. 1A depicts a block diagram of a computer system 10 suitable for implementing the present invention. Computer system 10 includes a bus 12 which interconnects major subsystems such as a central processor 14, a system memory 16 (typically RAM), an input/output (I/O) controller 18, an external device such as a display screen 24 via a display adapter 26, serial ports 28 and 30, a keyboard 32, a fixed disk drive 34 via a storage interface 35 and a floppy disk drive 36 operative to receive a floppy disk 38, and a CD-ROM (or DVD-ROM) device 40 operative to receive a CD-ROM 42. Many other devices can be connected such as a user pointing device, e.g., a mouse 44 connected via serial port 28 and a network interface 46 connected via serial port 30.

Many other devices or subsystems (not shown) may be connected in a similar manner. Also, it is not necessary for all of the devices shown in Fig. 1A to be present to practice the present invention, as discussed below. The devices and subsystems may be interconnected in different ways from that shown in Fig. 1A. The operation of a computer system such as that shown in Fig. 1A is well known. Databases storing polymorphism information according to the present invention can be stored, e.g., in system memory 16 or on storage media

such as fixed disk 34, floppy disk 38, or CD-ROM 42. An application program to access such databases can be operably disposed in system memory 16 or sorted on storage media such as fixed disk 34, floppy disk 38, or CD-ROM 42.

Fig. 1B depicts the interconnection of computer system 10 to remote computers 48, 50, and 52. Fig. 1B depicts a network 54 interconnecting remote servers 48, 50, and 52. Network interface 46 provides the connection from client computer system 10 to network 54. Network 54 can be, e.g., the Internet. Protocols for exchanging data via the Internet and other networks are well known. Information identifying the polymorphisms described herein can be transmitted across network 54 embedded in signals capable of traversing the physical media employed by network 54.

Information identifying polymorphisms shown in Table 1 is represented in records, which optionally, are subdivided into fields. Each record stores information relating to a different polymorphisms in Table 1. Collectively, the records can store information relating to all of the polymorphisms in Table 1, or any subset thereof, such as 5, 10, 50, or 100 polymorphisms from Table 1. In some databases, the information identifies a base occupying a polymorphic position and the location of the polymorphic position. The base can be represented as a single letter code (i.e., A, C, G or T/U) present in a polymorphic form other than that in the reference allele. Alternatively, the base occupying a polymorphic site can be represented in IUPAC ambiguity code as shown in Table 1. The location of a polymorphic site can be identified as its position within one of the sequences shown in Table 1. For example, in the first sequence shown in Table 1, the polymorphic site occupies the 16th base. The position can also be identified by reference to, for example, a chromosome, and distance from known markers within the chromosome. In other databases, information

identifying a polymorphism contains sequences of 10-100 bases shown in Table 1 or the complements thereof, including a polymorphic site. Preferably, such information records at least 10, 15, 20, or 30 contiguous bases of sequences including a polymorphic site.

EXAMPLES

The polymorphisms shown in Table 1 were identified by resequencing of target sequences from eight unrelated individuals of diverse ethnic and geographic backgrounds by hybridization to probes immobilized to microfabricated arrays. The strategy and principles for design and use of such arrays are generally described in WO 95/11995. The strategy provides arrays of probes for analysis of target sequences showing a high degree of sequence identity to the reference sequences of the fragments shown in Table 1, column 1. The reference sequences were sequence-tagged sites (STSs) developed in the course of the Human Genome Project (*see, e.g., Science* 270, 1945-1954 (1995); *Nature* 380, 152-154 (1996)). Most STS's ranged from 100 bp to 300 bp in size.

A typical probe array used in this analysis has two groups of four sets of probes that respectively tile both strands of a reference sequence. A first probe set comprises a plurality of probes exhibiting perfect complementarily with one of the reference sequences. Each probe in the first probe set has an interrogation position that corresponds to a nucleotide in the reference sequence. That is, the interrogation position is aligned with the corresponding nucleotide in the reference sequence, when the probe and reference sequence are aligned to maximize complementarily between the two. For each probe in the first set, there are three corresponding probes from three additional probe sets. Thus, there are four probes corresponding to each

nucleotide in the reference sequence. The probes from the three additional probe sets are identical to the corresponding probe from the first probe set except at the interrogation position, which occurs in the same position in each of the four corresponding probes from the four probe sets, and is occupied by a different nucleotide in the four probe sets. In the present analysis, probes were 25 nucleotides long. Arrays tiled for multiple different references sequences were included on the same substrate.

Multiple target sequences from an individual were amplified from human genomic DNA using primers for the fragments indicated in the listed Web sites. The amplified target sequences were fluorescently labelled during or after PCR. The labelled target sequences were hybridized with a substrate bearing immobilized arrays of probes. The amount of label bound to probes was measured. Analysis of the pattern of label revealed the nature and position of differences between the target and reference sequence. For example, comparison of the intensities of four corresponding probes reveals the identity of a corresponding nucleotide in the target sequences aligned with the interrogation position of the probes. The corresponding nucleotide is the complement of the nucleotide occupying the interrogation position of the probe showing the highest intensity (see WO 95/11995). The existence of a polymorphism is also manifested by differences in normalized hybridization intensities of probes flanking the polymorphism when the probes hybridized to corresponding targets from different individuals. For example, relative loss of hybridization intensity in a "footprint" of probes flanking a polymorphism signals a difference between the target and reference (i.e., a polymorphism) (see EP 717,113, incorporated by reference in its entirety for all purposes). Additionally, hybridization intensities for corresponding targets from different individuals can be

classified into groups or clusters suggested by the data, not defined *a priori*, such that isolates in a give cluster tend to be similar and isolates in different clusters tend to be dissimilar. See WO 97/29212 (incorporated by reference in its entirety for all purposes). Hybridizations to samples from different individuals were performed separately. Table 1 summarizes the data obtained for target sequences in comparison with a reference sequence for the eight individuals tested.

From the foregoing, it is apparent that the invention includes a number of general uses that can be expressed concisely as follows. The invention provides for the use of any of the nucleic acid segments described above in the diagnosis or monitoring of diseases, such as cancer, inflammation, heart disease, diseases of the CNS, and susceptibility to infection by microorganisms. The invention further provides for the use of any of the nucleic acid segments in the manufacture of a medicament for the treatment or prophylaxis of such diseases. The invention further provides for the use of any of the DNA segments as a pharmaceutical.

All publications and patent applications cited above are incorporated by reference in their entirety for all purposes to the same extent as if each individual publication or patent application were specifically and individually indicated to be so incorporated by reference. Although the present invention has been described in some detail by way of illustration and example for purposes of clarity and understanding, it will be apparent that certain changes and modifications may be practiced within the scope of the appended claims.

WHAT IS CLAIMED IS:

1 1 A nucleic acid segment of between 10 and 100
2 bases from a fragment shown in Table 1 including a
3 polymorphic site, or the complement of the segment.

1 2. The nucleic acid segment of claim 1 that is
2 DNA.

1 3. The nucleic acid segment of claim 1 that is
2 RNA.

1 4. The segment of claim 1 that is less than 50
2 bases.

1 5. The segment of claim 1 that is less than 20
2 bases.

1 6. The segment of claim 1, wherein the fragment
2 is WI-14263 and the polymorphic site is at position 49.

1 7. The segment of claim 1, wherein the
2 polymorphic site is diallelic.

1 8. The segment of claim 1, wherein the
2 polymorphic form occupying the polymorphic site is the
3 reference base for the fragment listed in Table 1, column
4 3.

1 9. The segment of claim 1, wherein the
2 polymorphic form occupying the polymorphic site is an
3 alternative form for the fragment listed in Table 1,
4 column 5.

1 10. An allele-specific oligonucleotide that
2 hybridizes to a segment of a fragment shown in Table 1,
3 column 8 or its complement.

11. The allele-specific oligonucleotide of claim 10 that is probe.

12. The allele-specific oligonucleotide of claim 10, wherein a central position of the probe aligns with the polymorphic site of the fragment.

13. The allele-specific oligonucleotide of claim 10 that is a primer.

14. The allele-specific oligonucleotide of claim 13, wherein the 3' end of the primer aligns with the polymorphic site of the fragment.

15. An isolated nucleic acid comprising a sequence of Table 1, column 8 or the complement thereof, wherein the polymorphic site within the sequence or complement is occupied by a base other than the reference base show in Table 1, column 3.

16. A method of analyzing a nucleic acid, comprising:
obtaining the nucleic acid from an individual; and
determining a base occupying any one of the polymorphic sites shown in Table 1.

17. The method of claim 16, wherein the determining comprises determining a set of bases occupying a set of the polymorphic sites shown in Table 1.

18. The method of claim 16, wherein the nucleic acid is obtained from a plurality of individuals, and a base occupying one of the polymorphic positions is determined in each of the individuals, and the method further comprising testing each individual for the

presence of a disease phenotype, and correlating the presence of the disease phenotype with the base.

19. A computer-readable storage medium for storing data for access by an application program being executed on a data processing system, comprising:

a data structure stored in the computer-readable storage medium, the data structure including information resident in a database used by the application program and including:

a plurality of records, each record of the plurality comprising information identifying a polymorphisms shown in Table 1.

20. The computer-readable storage medium of claim 19, wherein each record has a field identifying a base occupying a polymorphic site and a location of the polymorphic site.

21. The computer-readable storage medium of claim 19, wherein each record identifies a nucleic acid segment of between 10 and 100 bases from a fragment shown in Table 1 including a polymorphic site, or the complement of the segment.

22. The computer-readable storage medium of claim 19, comprising at least 10 records, each record comprising information identifying a different polymorphism shown in Table 1.

23. The computer-readable storage medium of claim 19, comprising at least 100 records, each record comprising information identifying a different polymorphisms shown in Table 1.

24. A signal carrying data for access by an application program being executed on a data processing system, comprising:

a data structure encoded in the signal, said data structure including information resident in a database used by the application program and including:

a plurality of records, each record of the plurality comprising information identifying a polymorphism shown in Table 1.

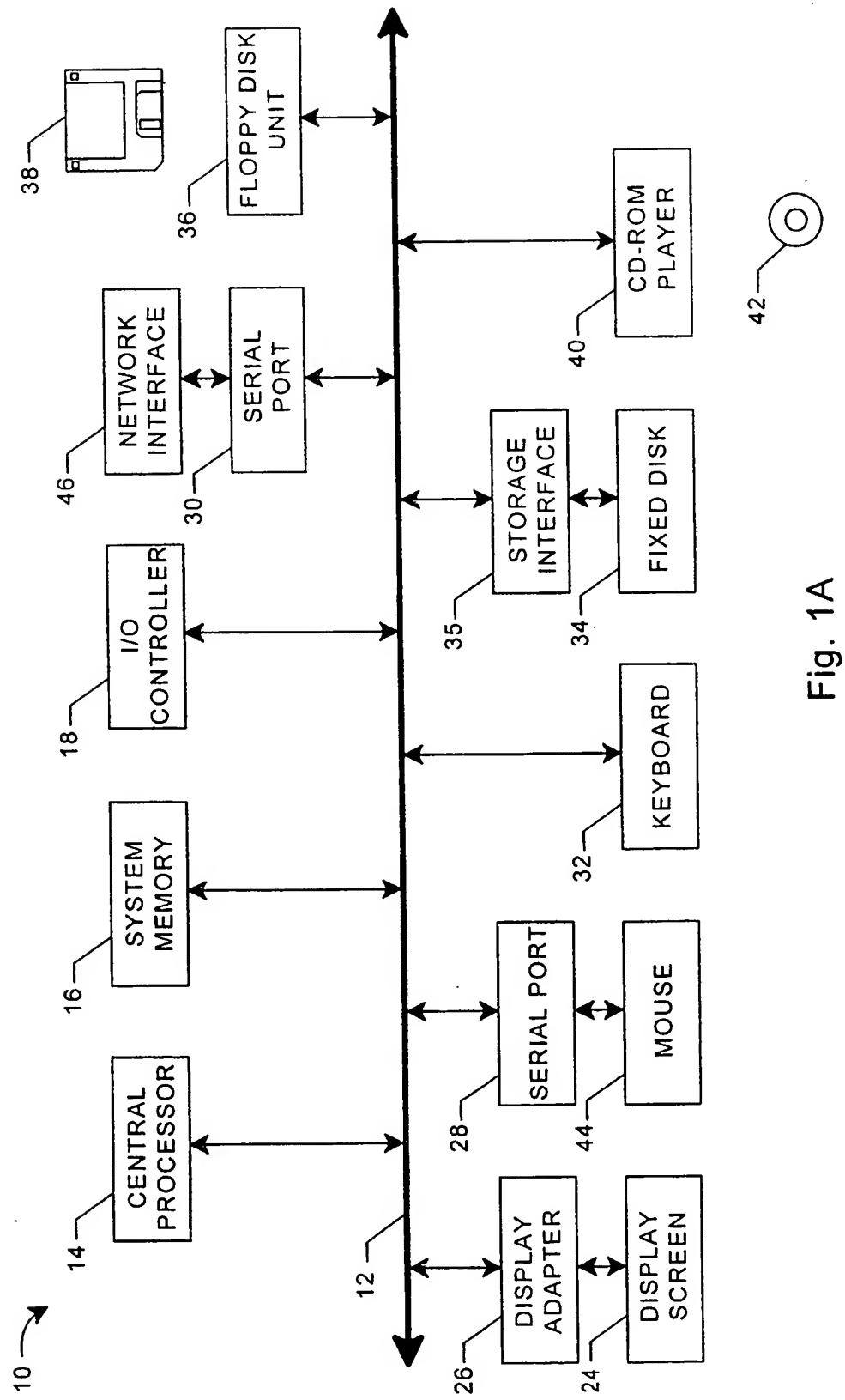
10
12
14
CENTRAL PROCESSOR
16
SYSTEM MEMORY
18
I/O CONTROLLER
24
DISPLAY SCREEN
26
DISPLAY ADAPTER
28
SERIAL PORT
30
SERIAL PORT
32
KEYBOARD
34
FIXED DISK
35
STORAGE INTERFACE
36
FLOPPY DISK UNIT
38
40
CD-ROM PLAYER
42
44
MOUSE
46
NETWORK INTERFACE

Fig. 1A

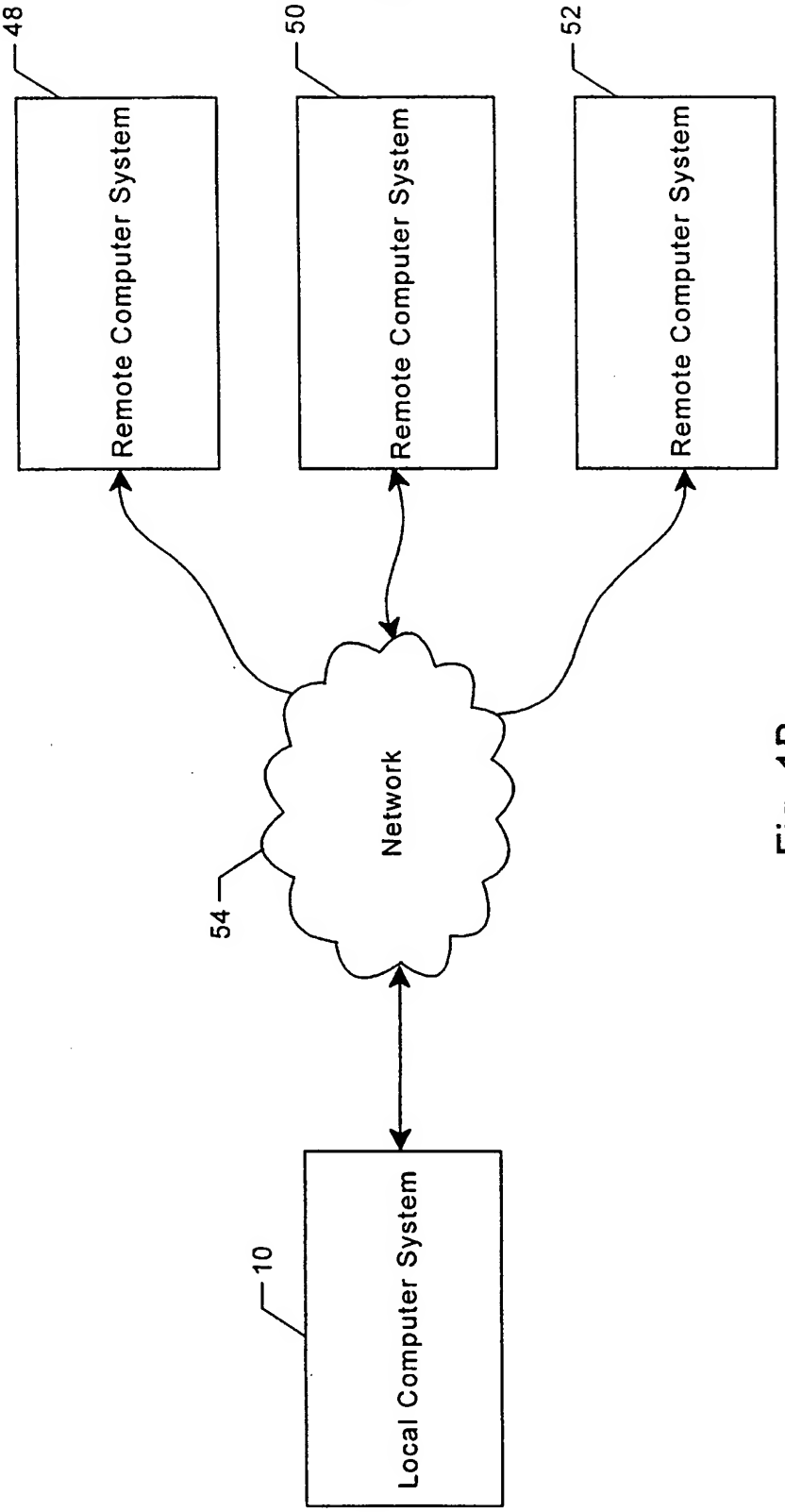
48
Remote Computer System
50
Remote Computer System
52
Remote Computer System
Network
54
10
Local Computer System

Fig. 1B

INTERNATIONAL SEARCH REPORT

International application No.
PCT/US98/19325

A. CLASSIFICATION OF SUBJECT MATTER
IPC(6) :C07H 21/00
US CL 536/23.1
According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)

U.S. : Please See Extra Sheet.

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)

Please See Extra Sheet.

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
A	MATTHEWS et al. Analytical Strategies for the Use of DNA Probes. Analytical Biochemistry. February 1988, Volume 169, pages 1-25, see the entire document.	1-24
Y	Sigma Chemical Catalog, (Published in 1990 by Sigma Chemical Company, P.O. Box 14508, Saint Louis, Missouri 63178) page 845, see especially Product P 0887 as compared to fragment stSG3590 on page 17 of the instant description.	1,2,4,5, 8,10,11, 13

[X] Further documents are listed in the continuation of Box C. [ ] See patent family annex.

*	Special categories of cited documents:	"T"	later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention
"A"	document defining the general state of the art which is not considered to be of particular relevance	"X"	document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone
"E"	earlier document published on or after the international filing date	"Y"	document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art
"L"	document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)	"&"	document member of the same patent family
"O"	document referring to an oral disclosure, use, exhibition or other means		
"P"	document published prior to the international filing date but later than the priority date claimed		

Date of the actual completion of the international search
04 JANUARY 1999

Date of mailing of the international search report
22 JAN 1999

Name and mailing address of the ISA/US
Commissioner of Patents and Trademarks
Box PCT
Washington, D.C. 20231
Facsimile No. (703) 305-3230

Authorized officer
ARDIN MARSCHEL
Telephone No. (703) 308-0196

Form PCT/ISA/210 (second sheet)(July 1992)*

INTERNATIONAL SEARCH REPORT

International application No.
PCT/US98/19325

C (Continuation). DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
Y	YE et al. Progression of Coronary Atherosclerosis Is Associated with a Common Genetic Variant of the Human Stromelysin-1 Promoter Which Results in Reduced Gene Expression. The Journal of Biological Chemistry. 31 May 1996, Volume 271, Number 22, pages 13055-13060, see especially the abstract and page 13056, second column, first full paragraph.	1-24
Y	US 5,639,607 A (DESNICK ET AL.) 17 June 1997, see especially the abstract.	1-24
Y	US 5,449,604 A (SCHELLENBERG ET AL.) 12 September 1995, see especially the abstract and Table 1 in columns 15-18.	1-24
Y	US 4,965,190 A (WOO ET AL.) 23 October 1990, see especially the abstract and Figures 2B and 3.	1-24
Y	US 5,494,794 A (WALLACE) 27 February 1996, see especially the abstract and Figure 8.	1-24
Y	US 5,400,249 A (SOLL ET AL.) 21 March 1995, see the entire disclosure.	19-24

INTERNATIONAL SEARCH REPORT

International application No.
PCT/US98/19325

B. FIELDS SEARCHED
Minimum documentation searched
Classification System: U.S.

341/1,50,126,137; 360/1,32,40,131,135; 130; 365/49,52; 435/6; 536/23.1,24.1,24.3,24.31,24.32,24.33; 935/77,78

B. FIELDS SEARCHED
Electronic data bases consulted (Name of data base and where practicable terms used):

APS, BIOSIS, MEDLINE, WPI, BIOTECH ABS, EMBASE search terms: nucleic acid, hybridize, polymorphic, probe, pattern, computer, disk, floppy, memory